(12)特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2004 年12 月23 日 (23.12.2004)

PCT

(10) 国際公開番号
WO 2004/111996 A1

(51) 国際特許分類[7]: G10L 11/02, 11/06

(21) 国際出願番号: PCT/JP2004/008051

(22) 国際出願日: 2004 年6 月3 日 (03.06.2004)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(30) 優先権データ:
特願2003-165946 2003 年6 月11 日 (11.06.2003) JP

(71) 出願人 (米国を除く全ての指定国について): 松下電器産業株式会社 (MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD.) [JP/JP]; 〒5718501 大阪府門真市大字門真1006番地 Osaka (JP).

(72) 発明者; および

(75) 発明者/出願人 (米国についてのみ): 鈴木 哲 (SUZUKI, Tetsu). 金森 丈郎 (KANAMORI, Takeo). 河村 岳 (KAWAMURA, Takashi).

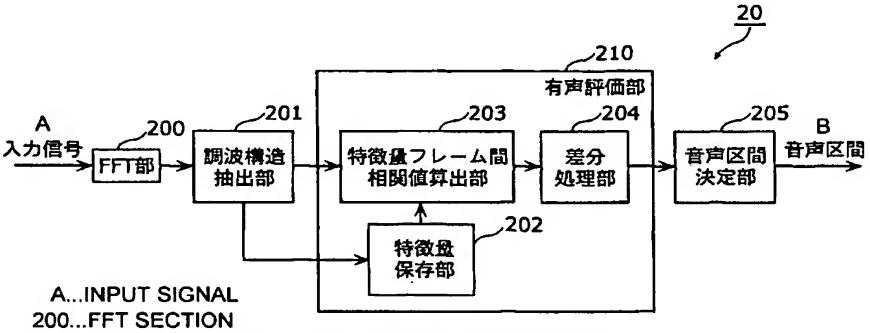
(74) 代理人: 新居 広守 (NII, Hiromori); 〒5320011 大阪府大阪市淀川区西中島3丁目11番26号 新大阪末広センタービル3F 新居国際特許事務所内 Osaka (JP).

(81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE,

[続葉有]

(54) Title: ACOUSTIC INTERVAL DETECTION METHOD AND DEVICE

(54) 発明の名称: 音響区間検出方法および装置



A...INPUT SIGNAL
200...FFT SECTION
201...HARMONIC STRUCTURE EXTRACTION SECTION
210...VOICED FEATURE EVALUATION SECTION
203...SECTION FOR CALCULATING CORRELATION VALUE BETWEEN FEATURE AMOUNT FRAMES
204...DIFFERENCE PROCESSING SECTION
202...FEATURE AMOUNT STORAGE SECTION
205...SPEECH SECTION DECISION SECTION
B...SPEECH INTERVAL

(57) Abstract: There is provided a harmonic-structured acoustic signal interval detection device not depending on the level fluctuation of the input signal, having an excellent real time characteristic, and noise resistance. The device includes: an FFT section (200) for subjecting the input signal to FFT and calculating a power spectrum component for each frame; a harmonic structure extraction section (201) for leaving only the harmonic structure from the power spectrum component; a voiced feature evaluation section (210) for evaluating correlation between the frames of harmonic structure extracted in the harmonic structure extraction section (201), thereby evaluating whether the interval is a vowel interval and extracting the voiced feature interval; and a speech interval decision section (205) for deciding a speech interval according to the continuity and durability of the output of the voiced feature evaluation section (210).

(57) 要約: 入力信号のレベル変動に依存せず、リアルタイム性に優れ、耐雑音性のある調波構造性音響信号区間検出装置は、入力信号にＦＦＴを施し、フレームごとにパワースペクトル成分を求めるＦＦＴ部（２００）と、パワースペクトル成分より調波構造のみを残す調波構造抽出部（２０１）

[続葉有]

WO 2004/111996 A1

SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:
— 国際調査報告書
— 補正書・説明書

２文字コード及び他の略語については、定期発行される各*PCT*ガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

と、調波構造抽出部（２０１）で抽出された調波構造のフレーム間での相関性を評価することにより、母音の区間であるか否かを評価し、有声区間を抽出する有声評価部（２１０）と、有声評価部（２１０）の出力の連続性および持続性に基づいて音声区間を決定する音声区間決定部（２０５）とを備える。

明　細　書

音響区間検出方法および装置

技術分野

本発明は、入力音響信号から調波構造を有する信号とくに音声が含まれる区間を音声区間として検出する調波構造信号区間および調波構造性音響信号区間検出方法に関し、特に、環境雑音下における調波構造信号および調波構造性音響信号区間検出方法に関する。

背景技術

人間の音声は、声帯の振動と発声器官の共振によって形成されており、音の大きさや音の高低を区別するために声帯を制御して振動の周波数を変化させたり，鼻や舌などの発声器官の位置つまり声道形状を変動させたりすることで、人はさまざまな音を発声していることが知られている。このように生成される音声を、音響信号として捕えると、その特徴は、周波数とともに緩やかに変化する成分である、スペクトル包絡と、短時間の周期的（有声母音などの場合）にまたは非周期的に変化する成分（子音や無声母音の場合）である、スペクトル微細構造から構成されていることが知られている。前者のスペクトル包絡成分が発声器官の共振特性を表しており、人間の喉や口の形をあらわす特徴量として用いられ、たとえば音声認識の特徴量としても用いられている。一方、後者のスペクトル微細構造は、音源の周期性を表しており、声帯の基本周期（ピッチ）、音の高低を表す特徴量として用いられている。音声信号のスペクトルは、これら２つの要素の積で表現されている。とくに母音部などにおいて、後者の基本周期およびその高調波成分をよく残している信号は、音声の

調波構造とも呼ばれている。

従来、入力音響信号から音声区間を検出する手法は、様々提案されている。それらを大きく分類すると、入力音響信号の帯域パワーやスペクトルの概形を示すスペクトル包絡などの振幅情報を用いて識別する方法（以下、「方法１」という。）、口映像を動画像解析することにより、その開閉を検出する方法（以下、「方法２」という。）、音声や雑音を表現する音響モデルと入力音響信号の音響特徴量とを比較することにより音声区間を検出する方法（以下、「方法３」という。）、および音声の調音器官の特徴である声道形状によって形成されるスペクトル包絡形状や声帯振動によって形成される調波構造に着目して音声区間を決定する方法（以下、「方法４」という。）などがある。

しかし、方法１では、もともと振幅情報だけで音声と雑音とを識別することが難しいという問題を含んでいる。このため、方法１では、音声区間と雑音区間とを仮定し、音声区間と雑音区間とを区別するために設定したしきい値を再学習することにより、音声区間の検出を行なっている。したがって、学習過程において雑音区間の振幅が音声区間の振幅に対して大きくなる（すなわち音声雑音比（以下、「ＳＮＲ」という。）がＯｄＢ程度まで低下する）と、雑音区間であるか音声区間であるかの仮定そのものの精度が性能に影響し、しきい値学習の精度が劣化してしまう。その結果として、音声区間検出の性能が劣化するという問題がある。

また、方法２では、例えば音入力を用いずに画像だけを用いて口が開いたことを検出するようにすれば、その音声区間検出推定精度は、ＳＮＲとは無関係に一定に保つことが可能である。しかし、画像解析処理は音声信号の解析処理に比べて、コストが高いことと、口がカメラの方向に向いていない場合には音声区間の検出ができないという問題がある。

さらに、方法３では、想定した環境雑音下での性能は確保されるもの

の、雑音を想定することそのものが難しいため、この方法を使用できる環境は限定的となってしまう。その場の雑音環境を学習する手法も提案されているが、振幅情報を利用する方法（方法１）と同様に、学習方法の精度に依存して性能が劣化するという問題もある。

一方、音声の調音器官の特徴である、声道形状によって形成されるスペクトル包絡形状や声帯振動によって形成される調波構造に着目して音声区間を決定する方法（方法４）も提案されてきた。

スペクトル包絡形状を利用した方法には、帯域パワー例えばケプストラムの連続性を評価する方法などがあるが、ＳＮＲが低下した状況では雑音のオフセット成分との区別がつきにくくなるため、性能が劣化する。

調波構造に着目した方法として、ピッチ検出法はその手法の一つであり、時間軸上の自己相関や高次ケフレンシーを抽出する方法、周波数軸上の自己相関を行なう方法等が提案されている。しかし、これらの方法は、対象とする信号が単一のピッチ（高調波の基本周波数）を持つ信号でない場合には音声区間の抽出が困難であり、環境雑音によって抽出誤りが発生し易い等の問題がある。

また、複数種類の音響信号が混在した音響信号から、人の音声や特定の楽器音等の調波構造を持った音響信号を強調したり、抑圧したり、分離抽出したりする技術が知られている。例えば音声信号に対しては、雑音と音声信号とが混在した音響信号から雑音のみを抑圧する雑音抑圧装置（たとえば、特開平９－１５３７６９号公報参照。）が、また音楽に対しては演奏に含まれる旋律の分離方法や除去方法（たとえば、特開平１１－１４３４６０号公報参照。）が、それぞれ提案されている。

しかし、特開平９－１５３７６９号公報に記載の方法では、入力信号の線形予測残差信号を帯域ごとに観察することで音声および非音声の検出を行っている。したがって、線形予測がうまく機能しない低ＳＮＲの

非定常雑音下では性能が劣化するという問題がある。

また、特開平１１－１４３４６０号公報に記載の方法は、同一の音程の音が一定時間持続するという音楽の旋律特有の性質を利用した方法である。このため、この方法を、音声と雑音との区別にそのまま用いることは困難であるという問題がある。音響の分離や除去を目的としない場合には、その処理量の多さが問題となる。

調波構造を表現する音響特徴量そのものを評価関数に用いる手法（たとえば、特開２００１—２２２２８９号公報参照。）も提案されている。図３２は、特開２００１—２２２２８９号公報で提案されている方法を用いた音声区間決定装置の概略構成を示すブロック図である。

図３２に示される音声区間検出装置１０は、入力信号中の音声区間を決定する装置であり、ＦＦＴ（Fast Fourier Transform）部１００と、調波構造評価部１０１と、調波構造ピーク検出部１０２と、ピッチ候補検出部１０３と、フレーム間振幅差分調波構造評価部１０４と、音声区間決定部１０５とを備える。

ＦＦＴ部１００は、入力信号に対し、フレーム（たとえば、１フレームは、１０ｍｓｅｃ）ごとにＦＦＴ処理を行ない、入力信号を周波数変換し、各種の分析を行なう。調波構造評価部１０１は、ＦＦＴ部１００より得られた周波数分析結果より、フレームごとに調波構造を有するか否かの評価を行なう。調波構造ピーク検出部１０２は、調波構造評価部１０１で抽出された調波構造をローカルピーク形状に変換し、ローカルピークを検出する。

ピッチ候補検出部１０３は、調波構造ピーク検出部１０２で検出されたローカルピークを時間軸方向（フレーム方向）にトラッキングすることによりピッチ検出を行なう。ピッチとは、調波構造の基本周波数のことである。

フレーム間振幅差分調波構造評価部１０４は、ＦＦＴ部１００における周波数分析の結果得られた振幅をフレーム間で差分し、差分値を求め、その差分値より着目しているフレームが調波構造を有するか否かの評価を行なう。

音声区間決定部１０５は、ピッチ候補検出部１０３で検出されたピッチと、フレーム間振幅差分調波構造評価部１０４の評価結果とを総合的に判断し、音声区間を決定する。

したがって、図３２に示される音声区間検出装置１０では、単一のピッチのみを有する音響信号のみならず、複数のピッチを有する音響信号であっても、音声区間を決定できる。

しかしながら、ピッチ候補検出部１０３において、ローカルピークをトラッキングする際には、ローカルピークの出現や消滅などを考慮しなければならず、これらを考慮しつつ、高精度でピッチを検出するのは困難である。

また、ピークという極大値を扱う性質上、雑音に対する耐性もあまり期待できない。さらに、時間的な変動を評価するために、フレーム間振幅差分調波構造評価部１０４においては、フレーム間差分に対して調波構造の有無を評価しているが、単に、振幅の差分を用いているため、調波構造の有する情報が失われてしまうだけではなく、例えば突発雑音が生じた場合には、差分値として突発雑音の音響特徴量がそのまま評価されてしまうという問題がある。

そこで、本発明は上述の課題を解決するためになされたものであり、入力信号のレベル変動に依存せず、精度良く音声区間を検出可能な調波構造性音響信号区間検出方法および装置を提供することを目的とする。

また、リアルタイム性に優れた調波構造性音響信号区間検出方法および装置を提供することも目的とする。

発明の開示

本発明のある局面に係る調波構造性音響信号区間検出方法は、入力音響信号から調波構造を有する信号とくに音声が含まれる区間を音声区間として検出する調波構造性音響信号区間検出方法であって、前記入力音響信号に対し、所定の時間で区切られたフレーム単位で音響特徴量を抽出する音響特徴量抽出ステップと、前記音響特徴量の持続性を評価し、評価結果に従って音声区間を決定する区間決定ステップとを含むことを特徴とする。

このように、音響特徴量の持続性を評価することにより、音声区間の決定を行なっている。このため、ローカルピークをトラッキングする従来の方法のようにローカルピークの出現や消滅など、入力信号のレベル変動を考慮する必要がなく、精度よく音声区間を決定することができる。

好ましくは、前記音響特徴量抽出ステップでは、前記入力音響信号に対しフレーム単位で周波数変換を行ない、前記周波数変換の結果より調波構造のみを強調し、前記音響特徴量を抽出することを特徴とする。

音声（特に母音）には、調波構造が見られる。このため、調波構造を強調した音響特徴量を用いて音声区間を決定することにより、さらに精度よく音声区間を決定することができる。

さらに好ましくは、前記音響特徴量抽出ステップでは、さらに、前記周波数変換の結果より調波構造を抽出し、当該調波構造を含む所定の帯域の周波数変換の結果を、前記音響特徴量とすることを特徴とする。

調波構造が保たれている帯域のみからなる音響特徴量を用いて音声区間を決定することにより、さらに精度よく音声区間を決定することができる。

さらに好ましくは、前記区間決定ステップでは、前記音響特徴量のフレーム間における相関値に基づいて、前記持続性を評価することを特徴

とする。

このように、調波構造の持続性をフレーム間の音響特徴量の相関値により評価している。このため、フレーム間での振幅差分を取り調波構造の持続性を評価する従来方法に比べ、調波構造の有する情報を残した評価が可能である。よって、短いフレームにわたる突発雑音が生じたような場合であっても、そのような突発雑音を音声区間として検出することがなくなり、精度よく音声区間を決定することができる。

さらに好ましくは、前記区間決定ステップは、前記音響特徴量の持続性を評価する評価値を算出する評価ステップと、前記評価値の時間的な連続性を評価し、評価結果に従って音声区間を決定する音声区間決定ステップとを含むことを特徴とする。

音声区間決定ステップでの処理は、実施の形態に述べるように、時間的に連続する有声区間（評価値のみから求められた音声区間）を連結して音声区間を検出する処理に相当する。このように、時間的に連続する有声区間を連結し、音声区間を決定することにより、母音に比べ調波構造性評価値が小さい子音をも音声区間と決定することができる。

さらに、調波構造を有する区間を、詳細に評価することにより、音声か非音声である音楽かどうかを判定することが可能である。調波構造を有すると判定されたフレームにおいて、フレーム内部で最大あるいは最小の調波構造性値が検出された帯域の番号指数を連続的に評価することで、その検出が可能である。

また、フレーム間における調波構造持続性評価値を用いて、調波構造があるとみなされた区間において、該評価値の分散を用いて、音声あるいは音楽など調波構造が持続した区間からの変移なのか、調波構造を持つ突発的なノイズなのかを判別することが可能である。

また、上記調波構造に関する特徴を有する区間以外の区間に対しては、

無音とみなせるほど入力信号が小さい区間あるいは調波構造を有しない非調波構造の区間を判定することができる。

また、実施の形態５で示すように、音入力しながらフレーム単位で調波構造性の判定を行なう方法を開示する。

さらに好ましくは、前記区間決定ステップは、さらに、所定数のフレームにわたる前記評価ステップにおいて算出される前記評価値と第１の所定しきい値との比較に基づいて、前記入力音響信号の音声雑音比を推定するステップと、推定された前記音声雑音比が第２の所定しきい値以上の場合には、前記評価ステップにおいて算出される前記評価値に基づいて前記音声区間を決定するステップとを含み、前記音声区間決定ステップでは、前記音声雑音比が前記第２の所定しきい値未満の場合に、前記評価値の時間的な連続性を評価し、評価結果に従って前記音声区間を決定することを特徴とする。

これにより、入力音響信号の推定音声雑音比が良好な場合には、音響特徴量の持続性を評価する評価値の時間的な連続性を評価し、前記音声区間を決定する処理を省略することができる。このため、リアルタイム性に優れた音声区間の検出が可能になる。

なお、本発明は、以上のような調波構造性音響信号区間検出方法として実現することができるだけでなく、そのステップを手段とする調波構造性音響信号区間検出装置として実現したり、調波構造性音響信号区間検出方法の各ステップをコンピュータに実行させるためのプログラムとして実現したりすることもできる。そのようなプログラムは、ＣＤ－ＲＯＭ等の記録媒体やインターネット等の伝送媒体を介して配信することができるのはいうまでもない。

以上のように、本発明に係る調波構造性音響信号区間検出方法および装置によると、音声区間と雑音区間との精度良い選別が可能となり、特

に、音声認識方法の前処理として本発明を適用することにより、音声認識率を向上させることができ、その実用的価値は極めて高い。また、IC（Integrated Circuit）レコーダなどに使用することにより音声区間のみを録音したりすることにより、記録容量の効率利用も可能である。

図面の簡単な説明

図１は、本発明の実施の形態１に係る音声区間検出装置のハードウェア構成を示すブロック図である。

図２は、実施の形態１に係る音声区間検出装置が実行する処理のフローチャートである。

図３は、調波構造抽出部による調波構造抽出処理のフローチャートである。

図４（ａ）～図４（ｆ）は、各フレームにおけるスペクトル成分から調波構造のみを残したスペクトル成分を抽出する過程を模式的に示す図である。

図５（ａ）～図５（ｆ）は、本発明による入力信号の変換の遷移を示す図である。

図６は、音声区間決定処理のフローチャートである。

図７は、本発明の実施の形態２に係る音声区間検出装置のハードウェア構成を示すブロック図である。

図８は、実施の形態２に係る音声区間検出装置が実行する処理のフローチャートである。

図９は、実施の形態３に係る音声区間検出装置のハードウェア構成を示すブロック図である。

図１０は、音声区間検出装置が実行する処理のフローチャートである。

図１１は、調波構造抽出処理を説明するための図である。

図１２は、調波構造抽出処理の詳細を示すフローチャートである。

図１３（ａ）は、入力信号のパワースペクトルを示す図である。図１３（ｂ）は、調波構造性値Ｒ（ｉ）を示す図である。図１３（ｃ）は帯域番号Ｎ（ｉ）を示す図である。図１３（ｄ）は重み付き帯域番号Ｎｅ（ｉ）を示す図である。図１３（ｅ）は補正調波構造性値Ｒ’（ｉ）を示す図である。

図１４（ａ）は、入力信号のパワースペクトルを示す図である。図１４（ｂ）は、調波構造性値Ｒ（ｉ）を示す図である。図１４（ｃ）は帯域番号Ｎ（ｉ）を示す図である。図１４（ｄ）は重み付き帯域番号Ｎｅ（ｉ）を示す図である。図１４（ｅ）は補正調波構造性値Ｒ’（ｉ）を示す図である。

図１５（ａ）は、入力信号のパワースペクトルを示す図である。図１５（ｂ）は、調波構造性値Ｒ（ｉ）を示す図である。図１５（ｃ）は帯域番号Ｎ（ｉ）を示す図である。図１５（ｄ）は重み付き帯域番号Ｎｅ（ｉ）を示す図である。図１５（ｅ）は補正調波構造性値Ｒ’（ｉ）を示す図である。

図１６（ａ）は、入力信号のパワースペクトルを示す図である。図１６（ｂ）は、調波構造性値Ｒ（ｉ）を示す図である。図１６（ｃ）は帯域番号Ｎ（ｉ）を示す図である。図１６（ｄ）は重み付き帯域番号Ｎｅ（ｉ）を示す図である。図１６（ｅ）は補正調波構造性値Ｒ’（ｉ）を示す図である。

図１７は、音声・音楽区間決定処理の詳細なフローチャートである。

図１８は、実施の形態４に係る音声区間検出装置のハードウェア構成を示すブロック図である。

図１９は、音声区間検出装置が実行する処理のフローチャートである。

図２０は、調波構造抽出処理の詳細を示すフローチャートである。

図２１は、音声区間決定処理の詳細を示すフローチャートである。

図２２（ａ）は入力信号のパワースペクトルを示す図である。図２２（ｂ）は調波構造性値Ｒ（ｉ）を示す図である。図２２（ｃ）は、重み付き分散Ｖｅ（ｉ）を示す図である。図２２（ｄ）は連結前の音声区間を示す図である。図２２（ｅ）は連結後の音声区間を示す図である。

図２３（ａ）は入力信号のパワースペクトルを示す図である。図２３（ｂ）は調波構造性値Ｒ（ｉ）を示す図である。図２３（ｃ）は、重み付き分散Ｖｅ（ｉ）を示す図である。図２３（ｄ）は連結前の音声区間を示す図である。図２３（ｅ）は連結後の音声区間を示す図である。

図２４は、調波構造抽出処理の他の一例を示すフローチャートである。

図２５（ａ）は入力信号を示す図である。図２５（ｂ）は入力信号のパワースペクトルを示す図である。図２５（ｃ）は調波構造性値Ｒ（ｉ）を示す図である。図２５（ｄ）は重み付き調波構造性値Ｒｅ（ｉ）を示す図である。図２５（ｅ）は補正調波構造性値Ｒ’（ｉ）を示す図である。

図２６（ａ）は入力信号を示す図である。図２６（ｂ）は入力信号のパワースペクトルを示す図である。図２６（ｃ）は調波構造性値Ｒ（ｉ）を示す図である。図２６（ｄ）は重み付き調波構造性値Ｒｅ（ｉ）を示す図である。図２６（ｅ）は補正調波構造性値Ｒ’（ｉ）を示す図である。

図２７は、実施の形態５に係る音声区間検出装置６０の構成を示すブロック図である。

図２８は、音声区間検出装置の実行する処理のフローチャートである。

図２９（ａ）～図２９（ｄ）は、調波構造性区間の連結を説明するための図である。

図３０は、調波構造性フレーム仮判定処理の詳細なフローチャートである。

図３１は、調波構造性区間確定処理の詳細なフローチャートである。

図３２は、従来の音声区間決定装置の概略のハードウェア構成を示す図である。

発明を実施するための最良の形態

（実施の形態１）

以下、図面を参照しながら本発明の実施の形態１に係る音声区間検出装置について説明する。図１は、本実施の形態に係る音声区間検出装置２０のハードウェア構成を示すブロック図である。

音声区間検出装置２０は、入力音響信号（以下、単に「入力信号」という。）の中から人間が発声している区間である音声区間を決定する装置であり、ＦＦＴ部２００と、調波構造抽出部２０１と、有声評価部２１０と、音声区間決定部２０５とを備える。

ＦＦＴ部２００は、入力信号にＦＦＴを施し、フレームごとにパワースペクトル成分を求める。ここで、１フレームあたりの時間は１０ｍｓｅｃとするが、この時間に限定されるものではない。

調波構造抽出部２０１は、ＦＦＴ部２００で抽出されたパワースペクトル成分から雑音成分等を取り除き、調波構造のみを残したパワースペクトル成分を抽出する。

有声評価部２１０は、調波構造抽出部２０１で抽出された調波構造のみを残したパワースペクトル成分のフレーム間での相関性を評価することにより、母音の区間であるか否かを評価し、有声区間を抽出する装置であり、特徴量保存部２０２と、特徴量フレーム間相関値算出部２０３と、差分処理部２０４とを備える。なお、調波構造は、母音の発声区間内のパワースペクトル分布において主に見られる性質であり、子音の発声区間内のパワースペクトル分布においては、母音ほどの調波構造は見られない。

特徴量保存部２０２は、調波構造抽出部２０１より出力されるパワースペクトルを所定数のフレーム分保存する。特徴量フレーム間相関値算出部２０３は、調波構造抽出部２０１より出力されるパワースペクトルと、特徴量保存部２０２に保存されている一定フレーム前のパワースペクトルとの相関値を算出する。差分処理部２０４は、特徴量フレーム間相関値算出部２０３で求められた相関値のある一定期間における平均値を求め、特徴量フレーム間相関値算出部２０３より出力される相関値から平均値を引き、相関値と平均値との平均差分による補正相関値を求める。

音声区間決定部２０５は、差分処理部２０４より出力される平均差分による補正相関値に基づいて、音声区間を決定する。

以上のように構成された音声区間検出装置２０の動作について以下に説明する。図２は、音声区間検出装置２０が実行する処理のフローチャートである。

ＦＦＴ部２００は、調波構造を抽出するために使用する音響特徴量として、入力信号にＦＦＴを施すことにより、パワースペクトル成分を求める（Ｓ２）。より具体的には、ＦＦＴ部２００は、入力信号を所定のサンプリング周波数Ｆｓ（たとえば、１１．０２５ｋＨｚ）でサンプリングし、１フレーム（たとえば、１０ｍｓｅｃ）ごとに、所定のポイント（たとえば、１フレームあたり１２８ポイント）でＦＦＴのスペクトル成分を求める。ＦＦＴ部２００は、各ポイントで求められたスペクトル成分を対数化することによりパワースペクトル成分を求める。以下、パワースペクトル成分を、適宜単にスペクトル成分と表記する。

次に、調波構造抽出部２０１は、ＦＦＴ部２００で抽出されたパワースペクトル成分から雑音成分等を取り除き、調波構造のみを残したパワースペクトル成分を抽出する（Ｓ４）。

ＦＦＴ部２００で算出されたパワースペクトル成分には、雑音によるオフセットや声道形状によって形成されるスペクトル包絡形状が含まれており、それぞれが時間変動を起こしている。このため、調波構造抽出部２０１は、これらの成分を取り除き、声帯振動によって形成される調波構造のみを残したパワースペクトル成分をとりだす。これにより、より効果的に有声区間検出が行なわれる。

調波構造抽出部２０１による処理（Ｓ４）を図３および図４を参照しながらより詳細に説明する。図３は、調波構造抽出部２０１による調波構造抽出処理のフローチャートであり、図４は、各フレームにおけるスペクトル成分から調波構造のみを残したスペクトル成分を抽出する過程を模式的に示す図である。

図４（ａ）に示されるように、調波構造抽出部２０１は、各フレームのスペクトル成分Ｓ（ｆ）より、その極大値をピークホールドした値Ｈｍａｘ（ｆ）を算出し（Ｓ２２）、スペクトル成分Ｓ（ｆ）の極小値をピークホールドした値Ｈｍｉｎ（ｆ）を算出する（Ｓ２４）。

図４（ｂ）に示されるように、調波構造抽出部２０１は、スペクトル成分Ｓ（ｆ）から極小値のピークホールド値Ｈｍｉｎ（ｆ）を引くことにより、スペクトル成分Ｓ（ｆ）に含まれるフロア成分を除去する（Ｓ２６）。これにより、雑音オフセット成分およびスペクトル包絡に起因する変動成分が除去される。

図４（ｃ）に示されるように、調波構造抽出部２０１は、極大値のピークホールド値Ｈｍａｘ（ｆ）と極小値のピークホールド値Ｈｍｉｎ（ｆ）との差分値を求め、ピーク変動量を算出する（Ｓ２８）。

図４（ｄ）に示されるように、調波構造抽出部２０１は、ピーク変動量を周波数方向に微分し、その変化量を算出する（Ｓ３０）。これは、調波構造成分を有する帯域では、ピーク変動量の変化が小さいという仮定

に基づいて、調波構造の検出を行なうことを目的としている。

図４（ｅ）に示されるように、調波構造抽出部２０１は、上記仮定が反映されるような重みＷ（ｆ）を算出する（Ｓ３２）。すなわち、調波構造抽出部２０１は、ピーク変動量の変化量の絶対値と所定のしきい値とを比較し、当該変化量の絶対値が所定のしきい値θ以下であれば重みＷ（ｆ）を１とし、所定のしきい値θ以上であれば当該変化量の絶対値の逆数を重みＷ（ｆ）とする。これにより、ピーク変動量の変化が大きい部分の重みを小さくし、ピーク変動量の変化が小さい部分の重みを大きくすることができる。

図４（ｆ）に示されるように、調波構造抽出部２０１は、フロア成分が除去されたスペクトル成分（Ｓ（ｆ）－Ｈｍｉｎ（ｆ））に重みＷ（ｆ）を掛け合わせ、スペクトル成分Ｓ’（ｆ）を求める（Ｓ３４）。この処理により、ピーク変動量の変化の大きい非調波構造成分を除去することが可能となる。

再度、図２に示される音声区間検出装置２０の動作説明を続ける。調波構造抽出処理（図２のＳ４、図３）の後、特徴量フレーム間相関値算出部２０３は、調波構造抽出部２０１より出力されるスペクトル成分と、特徴量保存部２０２に保存されている所定フレーム前のスペクトル成分との間の相関値を算出する（Ｓ６）。

ここでは、着目しているフレームをｊ番目のフレームとした場合、隣接するフレームのスペクトル成分を用いて相関値Ｅ１（ｊ）を求める方法について説明する。相関値Ｅ１（ｊ）は、次式（１）～（５）に従い求められる。すなわち、ｉフレームおよびｉ－１フレームの１２８ポイントにおけるパワースペクトル成分Ｐ（ｉ）およびＰ（ｉ－１）を次式（１）および（２）でそれぞれ表すものとする。また、パワースペクトル成分Ｐ（ｉ）およびＰ（ｉ－１）の相関関数ｘｃｏｒｒ（Ｐ（ｊ－１），

Ｐ（ｊ））の値を次式（３）で表すものとする。すなわち、相関関数ｘｃｏｒｒ（Ｐ（ｊ－１），Ｐ（ｊ））の値は、各ポイントにおける内積値からなるベクトル量である。ｚ１（ｉ）を次式（４）に示されるようにｘｃｏｒｒ（Ｐ（ｊ－１），Ｐ（ｊ））のベクトルの要素の最大値を求める。これをｊフレームの相関値Ｅ１（ｊ）としてもよいし、次式（５）で表されるようにたとえば３フレーム分加算した値を用いても良い。

$$P(i) = (p1(i), p2(i), \ldots, p128(i)) \quad \cdots (1)$$

$$P(i-1) = (p1(i-1), p2(i-1), \ldots, p128(i-1)) \quad \cdots (2)$$

$$xcorr(P(i-1), P(i)) = (p1(i-1) \times p1(i), p2(i-1) \times p2(i), \ldots, p128(i-1) \times p128(i)) \quad \cdots (3)$$

$$z1(i) = \max(xcorr(P(i-1), P(i))) \quad \cdots (4)$$

$$E1(j) = \sum_{i=j-2}^{j} z1(i) \quad \cdots (5)$$

相関値Ｅ１（ｊ）の一例を図５に示すグラフを用いて説明する。図５は、入力信号を処理することにより得られる信号を表すグラフである。図５（ａ）は入力信号の波形を示している。この波形は、掃除機の雑音（ＳＮＲ＝０．５ｄＢ）がある環境において、約１２００～３０００ｍｓｅｃの間に「アールアンドビーホテルヒガシニホン」と発音している場合の波形である。この入力信号には、約５００ｍｓｅｃの箇所に掃除機を動かした際の「カタッ」という突発音が含まれ、２８００ｍｓｅｃ頃に掃除機のモータの回転速度を弱から強に変更し、掃除機の音のレベルが大きくなっている。図５（ｂ）は、図５（ａ）に示される入力信号

にＦＦＴを施した場合のパワーを示しており、図５（ｃ）は、相関値算出処理（Ｓ６）で求められた相関値の遷移を示している。

ここで、相関値Ｅ１（ｊ）の算出は、以下に示すような知見に基づいて算出される。すなわち、フレーム間の音響特徴量の相関値は、時間的に連続するフレームにおいて調波構造が連続していることに基づいている。このため、この調波構造を時間的に近いフレーム同士で相関をとることで、有声検出が行なわれる。調波構造が時間的に持続するのは主に母音区間である。このため、母音区間では相関値は大きくなり、子音区間では母音区間よりも相関値は小さくなるものと想定される。このように、調波構造に着目しフレーム間でパワースペクトル成分の相関値をとることによって、非周期的な雑音区間においては、相関値が小さくなるものと考えられる。このため、有声区間がより際立って識別可能となる。

また、一般的な発話スピードにおいて母音区間の持続時間は５０～１５０ｍｓｅｃ（５～１５フレーム）と言われており、その持続時間内であれば、フレーム間の相関係数の値は隣接するフレームでなくとも高くなるものと想定できる。この仮定が正しければ、やはり非周期的な雑音の影響を受けにくい評価関数であるということがいえる。相関値Ｅ１（ｊ）を算出する際に、数フレームにわたる相関関数の値の和を用いているのは、突発的に生じる雑音の影響を除去するためと、母音であれば、上記のように５０～１５０ｍｓｅｃの持続時間があるという知見によるものである。従って、図５（ｃ）に示されるように、５０フレームの近傍で発声する突発音に対しては反応せずに、相関値は小さいままである。

次に、差分処理部２０４は、特徴量フレーム間相関値算出部２０３で算出された相関値の一定時間にわたる平均値を求め、各フレームにおける相関値から当該平均値を減算し、平均差分による補正相関値を求める（Ｓ８）。なぜならば、相関値から平均値を引くことにより、長時間にわ

たり生じている周期性の雑音の影響を取り除くことができると考えられるためである。ここでは、５秒程度の相関値の平均値を求めており、図５（ｃ）では、平均値を実線５０２で示している。すなわち、実線５０２よりも上の部分に相関値が存在する区間が上記平均差分による補正相関値が正の区間である。

次に、音声区間決定部２０５は、主に有音区間を検出する相関値Ｅ１（ｊ）の差分処理部２０４で算出された平均差分による補正相関値に基づいて、後述する、相関値による選別、区間の持続長、子音区間や促音区間を加味した区間の連結、の３つの区間補正方法に従い音声区間を決定する（Ｓ１０）。

ここで、音声区間決定部２０５による音声区間決定処理（図２のＳ１０）についてより詳細に説明する。図６は、一発声単位で音声区間決定する処理の詳細を示すフローチャートである。

まず、第一の区間の補正方法である相関値による区間の判定について述べる。音声区間決定部２０５は、着目しているフレームについて、差分処理部２０４で求められた補正相関値が所定のしきい値よりも大きいか否かを調べる（Ｓ４４）。たとえば、所定のしきい値を０とした場合には、図５（ｃ）に示される相関値が相関値の平均値（実線５０２）よりも大きいか否かを調べることと等価である。

補正相関値が所定のしきい値よりも大きい場合には(Ｓ４４でＹＥＳ)、当該着目フレームは音声フレームであると判断し（Ｓ４６）、補正相関値が所定のしきい値以下の場合には（Ｓ４４でＮＯ）、当該着目フレームは非音声フレームであると判断する（Ｓ４８）。以上の音声判断処理（Ｓ４４～Ｓ４８）を音声区間検出対象となっているすべてのフレームについて繰返す（Ｓ４２～Ｓ５０）。以上の処理により、図５（ｄ）に示されるようなグラフが得られ、音声フレームが連続する区間が有声区間として

検出される。

このように、補正相関値の値がしきい値以下である場合には、そのフレームを非音声フレームであると判断する。ただし、騒音のレベルの影響や、音響特徴量のさまざまな条件に応じて、検出区間において期待される補正相関値が異なる。このため、音声フレームと非音声（雑音）フレームとを区別するためのしきい値は、事前の実験を通じて適宜定め用いることも可能である。この処理により調波構造性を有する信号の選別基準を厳しくすることにより、平均差分を求めた時間長より短い、例えば５００ｍｓ程度の周期雑音を非音声フレームとすることが期待できる。

次に、第二の区間の補正方法である隣接有声区間の連結法について述べる。音声区間決定部２０５は、着目している有声区間と、当該有声区間に隣接する有声区間との間の距離が所定フレーム数未満であるかを調べる（Ｓ５４）。たとえば、ここでは所定フレーム数を３０フレームとする。当該距離が３０フレーム未満の場合には（Ｓ５４でＹＥＳ）、隣接する２つの有声区間を連結する（Ｓ５６）。以上の処理（Ｓ５４～Ｓ５６）をすべての有声区間について行なう（Ｓ５２～Ｓ５８）。以上の有声区間連結処理により、図５（ｅ）に示されるようなグラフが得られ、近接する有声区間が連結されていることが分かる。

有声区間の連結をするのは、以下のような理由による。すなわち、子音区間、特に破裂音（／ｋ／，／ｃ／，／ｔ／，／ｐ／）や摩擦音などの無声子音の区間においては、調波構造が表れにくいため、相関値が小さく、有声区間として検出されにくい。しかし、子音の近傍には母音が存在するため、母音が連続する区間は有声区間とみなされるという理由による。これにより、子音部分も有声区間とすることが可能になる。

最後に、第三の区間の補正方法である区間持続時間について述べる。音声区間決定部２０５は、着目している有声区間について、その持続時

間が所定時間よりも長いか否かを調べる（Ｓ６２）。たとえば、所定時間は、５０ｍｓｅｃであるとする。持続時間が５０ｍｓｅｃよりも長い場合には（Ｓ６２でＹＥＳ）、当該有声区間を音声区間と決定し（Ｓ６４）、持続時間が５０ｍｓｅｃ以下の場合には（Ｓ６２でＮＯ）、当該有声区間を非音声区間と決定する（Ｓ６６）。以上の処理（Ｓ６２～Ｓ６６）をすべての有声区間について行なうことにより音声区間が決定される（Ｓ６０～Ｓ６８）。以上説明した処理により、図５（ｆ）に示すようなグラフが得られ、１１０～２８０フレームあたりに音声区間が検出される。また、図５（ｅ）のグラフに存在していた３２５フレームあたりに存在していた周期性ノイズに対する有声区間は、非音声区間と決定されていることが分かる。このように、有声区間の持続時間により有声区間を選別する処理では、相関値が高い短時間の周期的雑音を取り除くことができる。

以上説明したように本実施の形態によれば、調波構造を有するスペクトル成分のフレーム間での持続性を評価することにより、有声区間を決定している。このため、ローカルピークをトラッキングする従来の方法に比べ、精度よく音声区間を決定することができる。

特に、調波構造の持続性をフレーム間のスペクトル成分の相関値により評価している。このため、フレーム間での振幅差分を取り調波構造の持続性を評価する従来方法に比べ、調波構造の有する情報を残した評価が可能である。よって、短いフレームにわたる突発雑音が生じたような場合であっても、突発雑音を有声区間として検出することがない。

また、時間的に隣接する有声区間を連結することにより音声区間と決定している。このため、母音に比べ調波構造が小さい子音をも音声区間と決定することが可能である。また、有声区間の持続時間を評価することにより、周期性を有する雑音を除去することが可能になる。

（実施の形態２）

以下、図面を参照しながら本発明の実施の形態２に係る音声区間検出装置について説明する。本実施の形態に係る音声区間検出装置では、入力信号のＳＮＲがよい場合には、フレーム間でのスペクトル成分の相関性のみから音声区間を決定する点が実施の形態１に係る音声区間検出装置とは異なる。

図７は、本実施の形態に係る音声区間検出装置３０のハードウェア構成を示すブロック図である。実施の形態１に係る音声区間検出装置２０と同一の構成要素については、同一の参照番号を付す。その名称および機能も同一であるため、適宜説明を省略する。なお、以下の実施の形態においても同様に適宜説明を省略する。

音声区間検出装置３０は、入力信号の中から人間が発声している区間である音声区間を決定する装置であり、ＦＦＴ部２００と、調波構造抽出部２０１と、有声評価部２１０と、ＳＮＲ推定部２０６と、音声区間決定部２０５とを備える。

有声評価部２１０は、有声区間を抽出する装置であり、特徴量保存部２０２と、特徴量フレーム間相関値算出部２０３と、差分処理部２０４とを備える。

ＳＮＲ推定部２０６は、差分処理部２０４より出力される平均差分による補正相関値に基づいて、入力信号のＳＮＲを推定する。ＳＮＲ推定部２０６は、ＳＮＲが悪いと推定される場合には、差分処理部２０４より出力される補正相関値を音声区間決定部２０５に出力し、ＳＮＲがよいと推定される場合には、音声区間決定部２０５への補正相関値の出力は行なわずに、差分処理部２０４より出力される補正相関値より音声区間を決定する。これは、入力信号のＳＮＲが良好な場合には、音声区間と非音声区間との相関値の差がはっきりとしているという性質があるた

めである。

次に、ＳＮＲ推定部２０６による入力信号のＳＮＲの推定方法について説明する。ＳＮＲ推定部２０６は、差分処理部２０４で求められる相関値の平均値が所定のしきい値未満の場合には、ＳＮＲが良好であると推定し、当該平均値が所定のしきい値以上の場合には、ＳＮＲが悪いと推定する。これは、以下のような理由に基づく。すなわち、相関値の平均値を、一発声の持続時間よりも十分に長い時間（たとえば、５秒間）にわたって求めると、ＳＮＲが良好な環境下においては、雑音区間における相関値が小さくなるため、相関値の平均値が小さくなる。これに対し、周期性の雑音を有するようなＳＮＲが悪い環境下においては、雑音区間における相関値が大きくなるため、相関値の平均値が大きくなる。このように、相関値の平均値とＳＮＲとが連動しているという性質を用いることにより、既に計算済みの一つのパラメータを評価するだけで簡単にＳＮＲを推定することが可能である。

以上のように構成された音声区間検出装置３０の動作について以下に説明する。図８は、音声区間検出装置３０が実行する処理のフローチャートである。

ＦＦＴ部２００によるＦＦＴ処理（Ｓ２）から差分処理部２０４による補正相関値算出処理（Ｓ８）までは、図２に示した実施の形態１における音声区間検出装置２０の動作と同様である。そのため、その詳細な説明はここでは繰返さない。

次に、ＳＮＲ推定部２０６は、上記方法に従い、入力信号のＳＮＲを推定する（Ｓ１２）。ＳＮＲが良好であると推定される場合には（Ｓ１４でＹＥＳ）、所定のしきい値を超える補正相関値を音声区間として決定する（Ｓ１６）。ＳＮＲが悪いと推定される場合には（Ｓ１４でＮＯ）、図２および図６を参照して説明した実施の形態１に係る音声区間決定部２

０５による音声区間決定処理（図２のＳ１０）と同様の処理を実行し、音声区間を決定する（Ｓ１０）。

以上説明したように、本実施の形態によると、実施の形態１に記載の効果に加え、入力信号のＳＮＲが良好な場合には、有声区間の連続性および持続時間による音声区間決定処理を行なう必要がなくなる。このため、リアルタイム性に優れた音声区間の検出が可能になる。

（実施の形態３）

以下、図面を参照しながら本発明の実施の形態３に係る音声区間検出装置について説明する。本実施の形態に係る音声区間検出装置では、調波構造性を有する音声区間を決定するのみならず、音声区間の中から特に、音楽と人間の音声とを識別することができる。

図９は、本実施の形態に係る音声区間検出装置４０のハードウェア構成を示すブロック図である。音声区間検出装置４０は、入力信号の中から人間が発声している区間である音声区間と、音楽の区間である音楽区間とを決定する装置であり、ＦＦＴ部２００と、調波構造抽出部４０１と、音声・音楽区間決定部４０２とを備える。

調波構造抽出部４０１は、ＦＦＴ部２００で抽出されたパワースペクトル成分に基づいて、調波構造性を示す値を出力する処理部である。音声・音楽区間決定部４０２は、差分処理部２０４より出力された調波構造性を示す値に基づいて、音声区間および音楽区間を決定する処理部である。

以上のように構成された音声区間検出装置４０の動作について以下に説明する。図１０は、音声区間検出装置４０が実行する処理のフローチャートである。

ＦＦＴ部２００は、調波構造を抽出するために使用する音響特徴量として、入力信号にＦＦＴを施すことにより、パワースペクトル成分を求

める（Ｓ２）。

次に、調波構造抽出部４０１は、ＦＦＴ部２００で抽出されたパワースペクトル成分から、調波構造性を示す値を抽出する（Ｓ８２）。調波構造抽出処理（Ｓ８２）については、後に詳述する。

調波構造抽出部４０１は、調波構造性を示す値に基づいて、音声区間および音楽区間を決定する（Ｓ８４）。音声・音楽区間決定処理（Ｓ８４）については、後に詳述する。

次に、上述した調波構造抽出処理（Ｓ８２）について、詳細に説明する。調波構造抽出処理（Ｓ８２）では、パワースペクトル成分を複数の帯域に分割した際に、帯域間の相関を取ることにより、調波構造性を示す値を求める。このような方法により調波構造性を示す値を求めるのは、以下のような理由による。すなわち、調波構造性は、その発生源である声帯振動における信号の影響がよく残されている帯域に見られると仮定すると、隣接帯域との間で、パワースペクトル成分の相関性が高いという推測が成立するからである。すなわち、図１１に示すように、横軸に示す各フレームにおいて、縦軸に示すパワースペクトル成分を複数の帯域（この図において、帯域数は８）に区切った場合には、調波構造性を有する帯域間（例えば、帯域６０８と帯域６０６との間）においては、相関性が高いが、調波構造性を有しない帯域間（例えば、帯域６０２と帯域６０４との間）においては、相関性が低い。

図１２は、調波構造抽出処理（Ｓ８２）の詳細を示すフローチャートである。調波構造抽出部４０１は、各フレームについて、上述のように、各帯域間で帯域間相関値Ｃ（ｉ，ｋ）を算出する（Ｓ９２）。帯域間相関値Ｃ（ｉ，ｋ）は次式（６）で表される。

$$C(i,k)=\max(Xcorr(P(i,L*(k-1)+1:L*k),P(i,L*k+1:L*(k+1)))) \quad \cdots (6)$$

ここで、P(i,x:y)はフレームiのパワースペクトルにおける周波数成分x：y（x以上、y以下）での、ベクトル列を示す。また、Lは帯域幅を示し、max（Xcorr（・））はベクトル列間の相関係数の最大値を示す。

調波構造性を有する帯域では、隣接帯域との相関性が高いため、帯域間相関値C（i，k）が大きな値を示す。逆に、調波構造性を有しない帯域では、隣接帯域との相関性が低いため、帯域間相関値C（i，k）が小さな値を示す。

なお、帯域間相関値C（i，j）は次式（7）により求めてもよい。

C(i,k)=max(Xcorr(P(i,L*(k-1)+1:L*k),P(i+1,L*k+1:L*(k+1))) … (7)

なお、式（6）は、帯域608および帯域606間、または帯域604および帯域602間のように、同一フレーム内の隣接する帯域間でのパワースペクトルの相関を示しているのに対し、式（7）は、帯域608および帯域610間のように、隣接するフレーム間であり、かつ隣接する帯域間でのパワースペクトルの相関を示している。式（7）のように、隣接フレーム間でも相関を取ることにより、帯域間の相関とフレーム間の相関とを同時に計算することができる。

さらに、帯域間相関値C（i，k）は次式（8）により求めてもよい。

C(i,k)=max(Xcorr(P(i,L*(k-1)+1:L*k),P(i,L*(k-1)+1:L*(k+1))) … (8)

式（8）は、隣接フレームの同一帯域間でのパワースペクトルの相関を示している。

次に、フレームiにおける調波構造性を示す調波構造性値R（i）および帯域番号N（i）の組［R（i），N（i）］を求める（S94）。［R（i），N（i）］は、次式（9）に従い表される。

[R（i），N（i）] ＝ [R１（i）－R２（i），
N１（i）－N２（i）] …（９）

ただし、R１（i），R２（i）は以下のようにあらわされる。

$$R_1(i) = \max_{k=1..L-1}(C(i,k)); \quad (10)$$

$$R_2(i) = \min_{k=1..L-1}(C(i,k)); \quad (11)$$

C：フレームiの帯域kにおける帯域調波性尺度

L：帯域数

また、N１（i）およびN２（i）は、C（i，k）が最大となる帯域番号および最小となる帯域番号をそれぞれ示す。式（９）に示される調波構造性値は、同一フレーム内での帯域間相関値の最大値から最小値を引くことにより求められる。このため、調波構造性のあるフレームではその値が大きくなり、調波構造性の無いフレームではその値が小さくなる。また、最大値から最小値を引くことにより、帯域間相関値を正規化している効果もある。このため、図２のＳ８の処理のように、平均相関値との差分処理を行なうことなく、１つのフレームにおいて正規化処理を行なうことができる。

次に、調波構造抽出部４０１は、帯域番号N（i）をその過去Xcフレームにおける分散で重み付けした補正帯域番号Nd（i）を算出する（Ｓ９６）。また、調波構造抽出部４０１は、補正帯域番号Nd（i）の過去Xcフレームにおける最大値Ne（i）を求める（Ｓ９８）。最大値Ne（i）を以下では重み付き帯域番号と称する。

補正帯域番号Nd（i）および重み付き帯域番号Ne（i）はXc＝５とした場合、以下の式により求められる。

$$Nd(i) = \underset{k=i-Xc:i}{median}(N(k)) - \underset{k=i-Xc:i}{var}(N(k)); \quad (12)$$

$$Ne(i) = \underset{k=i:i+Xc}{max}(Nd(k)); \quad (13)$$

Nd：分散で補正した帯域番号

Ne：分散で補正した帯域番号Ndの過去Xcフレームの最大値

Xc: 分散計算フレーム幅

調波構造性のない区間では、帯域番号Ｎ（ｉ）の分散が大きくなる。このため、補正帯域番号Ｎｄ（ｉ）の値が小さな値（例えば、負の値）になり、これに伴ない、重み付き帯域番号Ｎｅ（ｉ）も小さな値になる。

さらに、調波構造抽出部４０１は、調波構造性値Ｒ（ｉ）を重み付き帯域番号Ｎｅ（ｉ）で補正し、補正調波構造性値Ｒ’（ｉ）を算出する（Ｓ１００）。補正調波構造性値Ｒ’（ｉ）は、次式（１４）に従い求められる。なお、ここで用いる調波構造性値Ｒ（ｉ）は、Ｓ８で算出した値を用いてもよい。

Ｒ’（ｉ）＝Ｒ（ｉ）＊Ｎｅ（ｉ）　…（１４）

図１３～図１５は、上述の調波構造抽出処理（Ｓ８２）の実験結果を示す図である。

図１３は、掃除機のノイズがある環境下（ＳＮＲ＝１０ｄＢ）で人間が音声を発声している場合の実験結果を示す図である。４０フレーム近傍には、掃除機を動かした際の「カタッ」という突発音が発生しており、およそ２８０フレーム前後で、掃除機のモーターの回転速度を弱から強に変更したために、掃除機の音のレベルが大きくなり、周期性ノイズが発せられているものとする。また、人間は８０フレームあたりから２８０フレームあたりまでの間に音声を発声しているものとする。

図１３（ａ）は入力信号のパワースペクトルを示しており、図１３（ｂ）は調波構造性値Ｒ（ｉ）を示しており、図１３（ｃ）は帯域番号Ｎ（ｉ）を示しており、図１３（ｄ）は重み付き帯域番号Ｎｅ（ｉ）を示してお

り、図１３（ｅ）は補正調波構造性値Ｒ’（ｉ）を示している。なお、図１３（ｃ）に示す帯域番号は、図を見やすくするために実際の帯域番号に－１を掛けているため、０に近いほど周波数が小さい。

図１３（ｃ）に示すように、突発音や周期性ノイズが発生している部分（図中破線で囲った部分）では、帯域番号Ｎ（ｉ）の変動が大きくなっている。このため、図１３（ｄ）に示すように、その部分の重み付き帯域番号Ｎｅ（ｉ）は小さな値を示し、それに伴ない、図１３（ｅ）に示すように、補正調波構造性値も小さくなっている。

図１４は、掃除機のノイズがほとんどない環境下（ＳＮＲ＝４０ｄＢ）で、図１３と同じ音声を発生した場合の実験結果を示す図である。このような環境下においても図１３と同様に、調波構造性のない部分の補正調波構造性値Ｒ’（ｉ）は小さくなっている（図１４（ｅ））。

図１５は、ボーカルの無い音楽に対する実験結果を示す図である。音楽では和音が出力されるため調波構造性を有するが、ドラムによりビートを刻む区間などでは調波構造性を有しない。図１５（ａ）は入力信号のパワースペクトルを示しており、図１５（ｂ）は調波構造性値Ｒ（ｉ）を示しており、図１５（ｃ）は帯域番号Ｎ（ｉ）を示しており、図１５（ｄ）は重み付き帯域番号Ｎｅ（ｉ）を示しており、図１５（ｅ）は補正調波構造性値を示している。なお、図１５（ｃ）に示す帯域番号は、図１３（ｃ）と同じ理由により、０に近いほど周波数が小さい。図１５（ｃ）の破線で囲っている部分では、ドラムによりビートが刻まれることにより、調波構造性が失われている。尾のため、その部分では、図１５（ｄ）に示すように重み付き帯域番号Ｎｅ（ｉ）が小さくなっている。したがって、図１５（ｅ）に示すように重み付き調波構造性値Ｒ’（ｉ）も小さくなっている。また、無声区間においても同様に調波構造性値Ｒ’（ｉ）が小さくなっている。

なお、Ｓ９４の処理において、フレームｉにおける調波構造性を示す調波構造性値Ｒ（ｉ）および帯域番号Ｎ（ｉ）の組［Ｒ（ｉ），Ｎ（ｉ）］を次式（１５）に従い求めてもよい。

［Ｒ（ｉ），Ｎ（ｉ）］＝［Ｒ１（ｉ）－Ｒ２（ｉ），
Ｎ１（ｉ）－Ｎ２（ｉ）］　…（１５）

ただし、Ｒ１（ｉ），Ｒ２（ｉ）は以下のようにあらわされる。

$$R_1(i)=\sum_{k=1..NSP}(C(i,k)) \quad (16)$$

$$R_2(i)=\sum_{k=L-NSP..L-1}(C(i,k)) \quad (17)$$

Ｃ：フレームiの帯域kにおける帯域調波性尺度。

Ｌ：帯域数

ＮＳＰ：音声ピッチ周波数帯域と仮定する帯域数

また、Ｎ１（ｉ）およびＮ２（ｉ）は、Ｃ（ｉ，ｋ）が最大となる帯域番号および最小となる帯域番号をそれぞれ示す。

なお、Ｒ１（ｉ）またはＲ２（ｉ）を調波構造性値Ｒ（ｉ）としてもよい。

図１６は、式（１５）に従い重み付き調波構造性値Ｒ’（ｉ）を求めた実験結果である。図１６は、掃除機のノイズがかなりある環境下（ＳＮＲ＝０ｄＢ）で人間が音声を発生している場合の実験結果を示す図である。なお、人間が音声を発生するタイミング、掃除機の突発音および周期性ノイズの発生タイミングは、図１３に示したものと同じである。ここでは、式（１５）において、Ｌ＝１６、ＮＳＰ＝２としたときの値を示している。

この場合においても、人間が発声しているフレームの重み付き調波構造性値Ｒ’（ｉ）は大きい値を示し、突発音および周期性ノイズが発生しているフレームにおいては、重み付き調波構造性値Ｒ’（ｉ）は小さい値

を示している。

次に、音声・音楽区間決定処理（図１０のＳ８４）について詳細に説明する。図１７は、音声・音楽区間決定処理（図１０のＳ８４）の詳細なフローチャートである。

音声・音楽区間決定部４０２は、フレームｉについて、パワースペクトルＰ（ｉ）が所定の閾値Ｐｍｉｎよりも大きいか否かを調べる（Ｓ１１２）。所定の閾値Ｐｍｉｎ以下の場合には（Ｓ１１２でＮＯ）、そのフレームは無音のフレームであると判断する（Ｓ１２６）。パワースペクトルＰ（ｉ）が所定の閾値Ｐｍｉｎよりも大きい場合には（Ｓ１１２でＹＥＳ）、補正調波構造性値Ｒ’（ｉ）が所定の閾値Ｒｍｉｎよりも大きいか否かを判断する（Ｓ１１４）。

補正調波構造性値Ｒ’（ｉ）が所定の閾値Ｒｍｉｎ以下の場合には（Ｓ１１４でＮＯ）、フレームｉが調波構造性の無い音のフレームであると判断する（Ｓ１２４）。補正調波構造性値Ｒ’（ｉ）が所定の閾値Ｒｍｉｎよりも大きい場合には（Ｓ１１４でＹＥＳ）、音声・音楽区間決定部４０２は、重み付き帯域番号Ｎｅ（ｉ）の単位時間平均値ａｖｅ＿Ｎｅ（ｉ）を算出し（Ｓ１１６）、当該単位時間平均値ａｖｅ＿Ｎｅ（ｉ）が所定の閾値Ｎｅ＿ｍｉｎよりも大きいか否かを調べる（Ｓ１１８）。ここでａｖｅ＿Ｎｅ（ｉ）は以下の式に従い求められる。すなわち、フレームｉを含むｄフレーム（ここでは５０フレームとした）におけるＮｅ（ｉ）の平均値を示している。

$$ave\_Ne(i) = \underset{k=i-d:i}{\text{average}}(\text{Ne(i)}); \qquad (18)$$

ｄ：単位時間平均値を求 めるフレーム数

ａｖｅ＿Ｎｅ（ｉ）が所定の閾値Ｎｅ＿ｍｉｎよりも大きい場合には（Ｓ１１８でＹＥＳ）、音楽と判断し（Ｓ１２０）、それ以外の場合には

（Ｓ１１８でＮＯ）、人間の音声のような調波構造性を有する音であると判断する（Ｓ１２２）。以上の処理（Ｓ１１２～Ｓ１２６）をすべてのフレームについて繰り返す（Ｓ１１０～Ｓ１２８）。

なお、以上のようにａｖｅ＿Ｎｅ（ｉ）の大きさにより調波構造性を有する音の中から音楽と音声とを分離したのは以下のような考え方に基づく。すなわち、音楽も音声も信号そのものには調波構造性を有する音であるが、音声は、有声音と無声音とが繰り返し出現される音であることより、調波構造性値が有声音の部分では大きく、無声音の部分では小さくなり、それらが短い周期で交互に繰り返される。一方、音楽は、和音が連続的に出力されるため調波構造性を有する期間が比較的長い時間連続し、調波構造性値が大きい状態が一定する。したがって、調波構造性値が音楽ではあまり変動しないものの、音声では変動することを示している。換言すれば、重み付き帯域番号Ｎｅ（ｉ）の単位時間平均値ａｖｅ＿Ｎｅ（ｉ）は、音楽の方が音声よりも大きくなる。

なお、調波構造性値の時間的連続性に着目して音声と音楽とを判別するようにしてもよい。すなわち、単位時間内に調波構造性値が小さくなるフレーム数がどの程度あるかを調べるようにしてもよい。そのため、例えば、重み付き帯域番号Ｎｅ（ｉ）が単位時間あたり負になる個数を数えるようにしてもよい。単位時間（例えば、着目しているフレームｉを含む過去５０フレーム）のうち、重み付き帯域番号Ｎｅ（ｉ）が負になるフレーム数をＮｅ＿ｃｏｕｎｔ（ｉ）とした場合に、Ｓ１１６でａｖｅ＿Ｎｅ（ｉ）の代わりにＮｅ＿ｃｏｕｎｔ（ｉ）を算出し、Ｓ１１８でフレーム数Ｎｅ＿ｃｏｕｎｔ（ｉ）が所定の閾値よりも大きい場合に音声とし、小さい場合に音楽とするようにしてもよい。

以上説明したように、本実施の形態では、各フレームにおけるパワースペクトル成分を複数の帯域に区切り、帯域間で相関を取っている。こ

のため、声帯振動における信号の影響が良く残されている帯域を抽出することができ、調波構造を確実に抽出することができる。

また、調波構造の変動や、調波構造の連続性に基づいて調波構造を有する音が音楽であるのか音声であるのかを判定することができる。

（実施の形態４）

次に、図面を参照しながら本発明の実施の形態４に係る音声区間検出装置について説明する。本実施の形態にかかる音声区間検出装置では、調波構造性値の分散に基づいて調波構造を有する音声区間を決定する。

図１８は、本実施の形態に係る音声区間検出装置５０のハードウェア構成を示すブロック図である。音声区間検出装置５０は、入力信号の中から調波構造性を有する音声区間を検出する装置であり、ＦＦＴ部２００と、調波構造抽出部５０１と、ＳＮＲ推定部２０６と、音声区間決定部５０２とを備える。

調波構造抽出部５０１は、ＦＦＴ部２００より出力されたパワースペクトル成分に基づいて、調波構造性を示す値を出力する処理部である。音声区間決定部５０２は、調波構造性を示す値および推定されたＳＮＲに基づいて、音性区間を決定する処理部である。

以上のように構成された音声区間検出装置５０の動作について以下に説明する。図１９は、音声区間検出装置５０が実行する処理のフローチャートである。ＦＦＴ部２００は、調波構造を抽出するために使用する音響特徴量として、入力信号にＦＦＴを施すことにより、パワースペクトル成分を求める（Ｓ２）。

次に、調波構造抽出部５０１は、ＦＦＴ部２００で抽出されたパワースペクトル成分から、調波構造性を示す値を抽出する（Ｓ１４０）。調波構造処理（Ｓ１４０）については、後述する。

ＳＮＲ推定部２０６は、調波構造性を示す値に基づいて、入力信号の

ＳＮＲを推定する（Ｓ１２）。ＳＮＲの推定方法は、実施の形態２と同様である。このため、その詳細な説明はここでは繰り返さない。

音声区間決定部５０２は、調波構造性を示す値および推定されたＳＮＲに基づいて音声区間を決定する（Ｓ１４２）。音声区間決定処理（Ｓ１４２）については、後に詳述する。

本実施の形態では、有声音と無声音との間の遷移区間に対して評価を加えることにより、音声区間決定の制度を向上させる。図６に示した音声区間決定方法では、（１）音声区間間の距離が所定フレーム未満であれば、音声区間を連結し（Ｓ５２）、（２）連結後の音声区間の持続時間が所定時間以下であればその区間を非音声区間としていた（Ｓ６０）。すなわち、無声音に対しては、（１）の処理において、Ｓ４２において有声音と判断された音声の区間の間のフレームに対してなんら評価を行うことなく、（２）の処理により連結されることを暗に期待する方法である。

音声区間を詳細にみると、有声音、無声音および騒音（非音声区間）の遷移関係から次の３つのグループ（Ａグループ、ＢグループおよびＣグループ）に分類できるものと考えられる。

Ａグループは有声音のグループであり、有声音から有声音への遷移、騒音から有声音への遷移、有声音から騒音への遷移が考えられる。

Ｂグループは、有声音と無声音が混在する音のグループであり、有声音から無声音への遷移と、無声音から有声音への遷移が考えられる。

Ｃグループは非有声音のグループであり、無声音から無声音への遷移、無声音から騒音への遷移、騒音から無声音への遷移、騒音から騒音への遷移が考えられる。

Ａグループに含まれる音については、調波構造性を示す値の精度に依存して有音区間のみが決定されるものである。これに対して、Ｂグループに含まれる音については、有声区間の周辺での音の遷移を評価するこ

とができれば、無声音区間をも抽出することが期待できるものと考えられる。Cグループに含まれる音については、無声音区間だけを騒音下で抽出することは非常に難しいと考えられる。これは、騒音の性質が簡単には規定できないため、または、無声音の騒音に対するＳＮＲが悪い場合が多いためである。

したがって、本実施の形態では、Ａグループのみを抽出して音声区間を決定していた図６の方法に加えて、有声音と無声音との間の遷移を評価することにより、Ｂグループの音の抽出を行なう。このことにより、音声区間の決定精度を向上させることができるものと考える。また、無声音から有声音への遷移区間および有声音から無声音への遷移区間において、調波構造性を示す値は大から小および小から大へとそれぞれ大きく変化していると仮定できる。このため、調波構造性を示す値を用いて有音区間と判断された区間周辺について、調波構造性を示す値の分散に基づく尺度を用いることより、この調波構造性の値の変化を捉えることができる。ここで、調波構造性を示す値の分散を重み付き分散Ｖｅと呼ぶ。

次に、調波構造抽出処理（図１９のＳ１４０）について、詳細に説明する。図２０は、調波構造抽出処理（Ｓ１４０）の詳細を示すフローチャートである。

調波構造抽出部５０１は、各フレームについて、帯域間相関値Ｃ（ｉ，ｋ）を算出する（Ｓ１５０）。帯域間相関値Ｃ（ｉ，ｋ）の算出は、図１２のＳ９２と同様である。このため、その詳細な説明はここでは繰り返さない。

次に、調波構造抽出部５０１は、帯域間相関値Ｃ（ｉ，ｋ）を用いて重み付き分散Ｖｅ（ｉ）を次式に従い算出する（Ｓ１５２）。

$$Ve(i) = \underset{k=1:L}{\mathrm{count}}\left(\text{if} \underset{j=i-Xc:i}{\mathrm{var}}(C(j,k)) > th\_var\_change \right) \quad (19)$$

ここで、Ｘｃ：フレーム幅（＝１６）

Ｌ：帯域数（＝１６）

ｔｈ＿ｖａｒ＿ｃｈａｎｇｅ：閾値

である。

また、関数ｖａｒ（）は括弧内の値の分散を示す関数であり、関数ｃｏｕｎｔ（）は、カッコ内の条件を満たす個数をカウントする関数であるものとする。

最後に、調波構造抽出部５０１は、調波構造性値Ｒ（ｉ）を算出する（Ｓ１５４）。この算出方法は、図１２のＳ９４と同様である。このため、その詳細な説明はここでは繰り返さない。

次に、図２１を参照して、音声区間決定処理（図１９のＳ１４２）について説明する。音声区間決定部５０２は、フレームｉについてＲ（ｉ）が閾値Ｔｈ＿Ｒより大きくかつＶｅ（ｉ）が閾値Ｔｈ＿Ｖｅより大きいか否かを判断する（Ｓ１８２）。上述の条件を満たす場合には（Ｓ１８２でＹＥＳ）、音声区間決定部５０２は、フレームｉを音声フレームであると判断し、満たさない場合には（Ｓ１８２でＮＯ）、非音声フレームであると判断する（Ｓ１８６）。音声区間決定部５０２は、以上の処理をすべてのフレームについて行なう（Ｓ１８０～Ｓ１８８）。次に、音声区間決定部５０２は、ＳＮＲ推定部２０６で推定されたＳＮＲが悪いか否かを判断し（Ｓ１９０）、推定ＳＮＲが悪い場合には、ループＢおよびループＣの処理を実行する（Ｓ５２～Ｓ６８）。ループＢおよびループＣの処理は図６に示したものと同様である。このため、その詳細な説明はここでは繰り返さない。

なお、推定ＳＮＲがよい場合には（Ｓ１９０でＮＯ）、ループＢを省略

し、ループＣの処理（Ｓ６０～Ｓ６８）のみを実行する。

図２２および図２３は、音声区間検出装置５０の実行する処理の結果を示す図である。図２２は、掃除機のノイズがある環境下（ＳＮＲ＝１０ｄＢ）で人間が音声を発声している場合の実験結果を示す図である。４０フレーム近傍は、掃除機を動かした際の「カタッ」という突発音が発生しており、およそ２８０フレーム前後で、掃除機のモーターの回転速度を弱から強に変更したために、掃除機の音のレベルが大きくなり、周期性ノイズが発せられているものとする。また、人間は８０フレームあたりから２８０フレームあたりまでの間に音声を発声しているものとする。

図２２（ａ）は入力信号のパワースペクトルを示しており、図２２（ｂ）は調波構造性値Ｒ（ｉ）を示しており、図２２（ｃ）は、重み付き分散Ｖｅ（ｉ）を示しており、図２２（ｄ）は連結前の音声区間を示しており、図２２（ｅ）は連結後の音声区間を示している。

図２２（ｄ）において、実線は、調波構造性値Ｒ（ｉ）を閾値処理（図６のループＡ（Ｓ４２～Ｓ５０））することにより得られる音声区間を示しており、破線は、調波構造性値Ｒ（ｉ）および重み付き分散Ｖｅ（ｉ）を閾値処理（図２１のループＡ（Ｓ１８０～Ｓ１８８））することにより得られる音声区間を示している。また、図２２（ｅ）において、破線は区間連結処理（図２１のＳ１９０～Ｓ６８）に従い、図２２（ｄ）の破線で示した音声区間を連結した後の処理結果を示しており、実線は区間連結処理（図６のＳ５２～Ｓ６８）に従い、図２２（ｄ）の実線で示した音声区間を連結した後の処理結果を示している。図２２（ｅ）に示されるように、重み付き分散Ｖｅ（ｉ）を用いることにより、正確に音声区間を抽出することができている。

図２３は、掃除機のノイズがほとんどない環境下（ＳＮＲ＝４０ｄＢ）

で、図２２と同じ音声を発生した場合の実験結果を示す図である。図２３（ａ）～図２３（ｅ）のグラフの意味は、図２２（ａ）～図２２（ｅ）のグラフの意味と同様である。図２３から、区間連結前の図２３（ｄ）と区間連結後の図２３（ｅ）とを比較すると、図２３（ｄ）の破線で示されるＳ１８０の結果は、図２３（ｅ）の実線と同様に音声区間が精度良く連結されていることを示している。したがって、推定ＳＮＲが非常によい場合には、図２１のＳ１９０の判定処理により、Ｓ５２～Ｓ５８の処理を行なわずに、音声区間が決定されても音声区間の検出性能を維持することが可能である。

以上説明したように、本実施の形態によると、重み付き分散Ｖｅを用いて無声音と有声音との遷移区間を評価することにより、上述のＢグループに属する音を抽出することができるようになった。このため、推定ＳＮＲを用いてＳＮＲがよいと判断された場合には区間連結を行わずとも音声区間が正確に抽出できるようになった。また、ＳＮＲが悪く、区間連結が必要な場合であっても、連結時の所定フレーム数（図２１のＳ５４）を小さくすることができるため、ノイズ区間を音声区間として誤検出することが少なくなった。

なお、以下に示すように調波構造性値Ｒ（ｉ）の代わりに補正調波構造性値Ｒ’（ｉ）を算出し、重み付き分散Ｖｅ（ｉ）と補正調波構造性値Ｒ’（ｉ）とから音声区間を検出するようにしてもよい。図２４は、調波構造抽出処理（図１９のＳ１４０）の他の一例を示すフローチャートである。

調波構造抽出部５０１は、帯域間相関値Ｃ（ｉ，ｋ）、重み付き分散Ｖｅ（ｉ）および調波構造性値Ｒ（ｉ）を算出する（Ｓ１６０～Ｓ１６４）。これらの算出方法は、図２０と同様であるため、その詳細な説明はここでは繰り返さない。次に、調波構造抽出部５０１は、重み付き調波構造

性値Ｒｅ（ｉ）を算出する（Ｓ１６６）。重み付き調波構造性値Ｒｅ（ｉ）は、次式に従い算出される。これらの式とＳ９６／Ｓ９８において算出される式との違いは、Ｓ９４において算出されるフレームｉにおける調波構造性値Ｒ（ｉ）を用いるかその帯域番号Ｎ（ｉ）を用いるかの違いにある。これらの式は、ともに、重み付き分散により補正されることにより、調波構造性を強調する指標となる。

$$Rd(i) = \underset{k=i-Xc:i}{median}(R(k)) - \underset{k=i-Xc:i}{var}(R(k)); \quad (20)$$

$$Re(i) = \underset{k=i:i+Xc}{max}(Rd(k)); \quad (21)$$

Xc: 分散計算フレーム幅 （＝5）

ここで、関数ｍｅｄｉａｎ（）は、括弧内の中央値を示す。

調波構造抽出部５０１は、補正調波構造性値Ｒ’（ｉ）を算出する（Ｓ１６８）。補正調波構造性値Ｒ’（ｉ）は以下の式に従い算出される。

$$R'(i) = Re(i); \ : \text{if } Re(i) > 0; \quad (22)$$

$$R'(i) = 0; \quad : \text{if } Re(i) < 0; \quad (23)$$

図２５および図２６は、図２４に示したフローチャートに従い処理された処理結果を示す図である。図２５は、掃除機のノイズが無い環境下（ＳＮＲ＝４０ｄＢ）で人間が音声を発声している場合の実験結果を示しており、図２６は、掃除機のノイズがある状況下（ＳＮＲ＝１０ｄＢ）で人間が音声を発声している場合の実験結果を示している。この実験では、図２３と同じ音声を発生するものとし、突発音と周期性ノイズの発生タイミングも同じであるものとする。

図２５（ａ）は入力信号を示し、図２５（ｂ）は入力信号のパワースペクトルを示しており、図２５（ｃ）は調波構造性値Ｒ（ｉ）を示して

おり、図２５（ｄ）は重み付き調波構造性値Ｒｅ（ｉ）を示しており、図２５（ｅ）は補正調波構造性値Ｒ’（ｉ）を示している。図２６（ａ）～図２６（ｅ）も図２５（ａ）～図２５（ｅ）とそれぞれ同様のグラフを示している。

補正調波構造性値Ｒ’（ｉ）は、調波構造性値Ｒ（ｉ）自身の分散に基づいて算出されている。このため、調波構造性を有する部分には当該分散が大きく、調波構造性を有しない部分では当該分散が小さいという性質を利用して、調波構造性を有する部分を適切に抽出することができる。

（実施の形態５）

上述した実施の形態１～４に記載の音声区間決定装置では、入力信号が予めファイル等に記録されている音声に対して区間決定を行なうものである。このような処理方法は、例えば、録音済みのデータに対して処理を行なう際には、有効であるが、音声を入力しながら区間決定を行なうには不向きである。そこで、本実施の形態においては、音声の入力に同期しながら音声区間をリアルタイムで決定する音声区間決定装置について説明する。

図２７は、本発明の実施の形態に係る音声区間検出装置６０の構成を示すブロック図である。音声区間検出装置６０は、入力信号から調波構造性を有する音声区間（調波構造性区間）を検出する装置であり、ＦＦＴ部２００と、調波構造抽出部６０１と、調波構造性区間確定部６０２と、制御部６０３とを備えている。

図２８は、音声区間検出装置６０の実行する処理のフローチャートである。制御部６０３は、ＦＲ、ＦＲＳ、ＦＲＥ、ＲＨ、ＲＭ、ＣＨ、ＣＭおよびＣＮを０にセットする（Ｓ２００）。ここで、ＦＲは、後述する調波構造性値Ｒ（ｉ）を未算出のフレームの先頭フレーム番号を示す。また、ＦＲＳは、調波構造性区間か否かが未確定の区間の先頭フレーム

番号を示す。ＦＲＥは、後述する調波構造性フレーム仮判定処理を行なった最終フレームのフレーム番号を示す。ＲＨおよびＲＭは調波構造性値の累積値を示す。ＣＨ、ＣＭおよびＣＮはカウンタである。

ＦＦＴ部２００は、入力フレームをＦＦＴ変換する。調波構造抽出部６０１は、ＦＦＴ部２００で抽出されたパワースペクトル成分に基づいて、調波構造性値Ｒ（ｉ）を抽出する。以上の処理を開始フレームＦＲから現在時刻のフレームＦＲＮまで行なう（Ｓ２０２～Ｓ２１０、ループＡ）。ループ処理が１回実行されるごとに、カウンタｉが１つずつインクリメントされ、開始フレームＦＲにカウンタｉの値が代入される（Ｓ２１０）。

次に、調波構造性区間確定部６０２は、ここまでで求められた調波構造性値Ｒ（ｉ）に基づいて、調波構造性を有する区間を仮判定する調波構造性フレーム仮判定処理を実行する（Ｓ２１２）。調波構造性フレーム仮判定処理については後述する。

調波構造性区間確定部６０２は、Ｓ２１２の処理の後、隣接する調波構造性区間が見つかったか否か、すなわち非調波構造性区間長ＣＮが０より大きいか否かを調べる（Ｓ２１４）。非調波構造性区間長ＣＮは、図２９（ａ）に図示するように、調波構造性区間の最終フレームと次の調波構造性区間の開始フレームとの間のフレーム長を示す。

隣接する調波構造性区間が見つかった場合には、非調波構造性区間長ＣＮが所定の閾値よりも小さいか否かを調べる（Ｓ２１６）。非調波構造性区間長ＣＮが所定の閾値ＴＨよりも小さければ（Ｓ２１６でＹＥＳ）、調波構造性区間確定部６０２は、図２９（ｂ）に示すように調波構造性区間を連結し、フレームＦＲＳ２からフレーム（ＦＲＳ２＋ＣＮ）までを調波構造性区間であると仮判定する（Ｓ２１８）。ここで、ＦＲＳ２とは、非調波構造性区間であると仮判定された最初のフレーム番号を示す。

非調波構造性区間長ＣＮが所定の閾値ＴＨ以上の場合には（Ｓ２１６でＮＯ）、図２９（ｃ）に示されるように調波構造性区間は連結されることなく、調波構造性区間確定部６０２が、後述する調波構造性区間確定処理を実行する（Ｓ２２０）。その後、制御部６０３は、ＦＳＲにＦＲＥを代入し、ＲＨ、Ｒｍ、ＣＨ、ＣＭおよびＣＮに０を代入する（Ｓ２２２）。調波構造性区間確定処理（Ｓ２２０）については後述する。

隣接する調波構造性区間が見つからなかった場合（Ｓ２１４でＮＯ、図２９（ｄ））、Ｓ２１８の処理の後、またはＳ２２２の処理の後、制御部６０３は、音声信号の入力が終了したか否かを判断する（Ｓ２２４）。音声信号の入力が終了していなければ（Ｓ２２４でＮＯ）、Ｓ２０２以降の処理が繰り返される。音声信号の入力が終了していれば（Ｓ２２４でＹＥＳ）、調波構造性区間確定部６０２は、調波構造性区間確定処理（Ｓ２２６）を実行し、処理を終了する。調波構造性区間確定処理（Ｓ２２６）については、後述する。

次に、調波構造性フレーム仮判定処理（図２８のＳ２１２）について説明する。図３０は、調波構造性フレーム仮判定処理の詳細なフローチャートである。調波構造性区間確定部６０２は、調波構造性値Ｒ（ｉ）が予め定められた調波構造性閾値１よりも大きいか否かを判断し（Ｓ２３２）、大きい場合には（Ｓ２３２でＹＥＳ）、着目しているフレームｉを調波構造性を有するフレームであると仮判断する。そして、累積調波構造性値ＲＨに調波構造性値Ｒ（ｉ）を加算し、カウンタＣＨを１つインクリメントする（Ｓ２３４）。

次に、調波構造性区間確定部６０２は、調波構造性値Ｒ（ｉ）が調波構造性閾値２よりも大きいか否かを判断し（Ｓ２３６）、大きい場合には（Ｓ２３６でＹＥＳ）、着目しているフレームｉを調波構造性を有する音楽のフレームであると仮判断する。そして、累積音楽調波構造性値ＲＭ

に調波構造性値Ｒ（ｉ）を加算し、カウンタＣＭを１つインクリメントする（Ｓ２３６）。以上の処理をフレームＦＲＥからフレームＦＲＮまで繰り返す（Ｓ２３０～Ｓ２３８）。

次に、調波構造性区間確定部６０２は、フレームＦＲＳ２をフレームＦＲＳとした後に、着目しているフレームｉの調波構造性値Ｒ（ｉ）が調波構造性閾値１よりも大きいか否かを判断し（Ｓ２４２）、大きい場合にはフレームＦＲＳ２をフレームｉとする（Ｓ２４４）。以上の処理をフレームＦＲＳからフレームＦＲＮまで繰り返す（Ｓ２４０～Ｓ２４６）。

次に、調波構造性区間確定部６０２は、カウンタＣＮを０にセットした後に、着目しているフレームｉの調波構造性値Ｒ（ｉ）が調波構造性閾値１以下であるか否かを判断し（Ｓ２５０）、調波構造性閾値１以下である場合には（Ｓ２５０でＹＥＳ）、フレームｉを非調波構造性区間であると仮判断し、カウンタＣＮを１つインクリメントする（Ｓ２５２）。以上の処理をフレームＦＲＳ２からフレームＦＲＮまで繰り返す（Ｓ２４８～Ｓ２５４）。以上の処理により、調波構造性を有する区間、音楽の調波構造性を有する区間および非調波構造性区間が仮判断される。

次に、調波構造性区間確定処理（図２８のＳ２２０、Ｓ２２６）について詳細に説明する。図３１は、調波構造性区間確定処理（図２８のＳ２２０、Ｓ２２６）の詳細なフローチャートである。

調波構造性区間確定部６０２は、調波構造性を有するフレーム数を示したカウンタＣＨの値が調波構造性フレーム長閾値１より大きく、かつ累積調波構造性値ＲＨが（ＦＲＳ－ＦＲＥ）×調波構造性閾値３よりも大きいか否かを判断する（Ｓ２６０）。上記条件を満たす場合には（Ｓ２６０でＹＥＳ）、フレームＦＲＳからフレームＦＲＥまでを調波構造性フレームであると判断する（Ｓ２６２）。

調波構造性区間確定部６０２は、音楽調波構造性を有するフレーム数

を示したカウンタＣＭの値が調波構造性フレーム長閾値２より大きく、かつ累積音楽調波構造性値ＲＭが（ＦＲＳ－ＦＲＥ）×調波構造性閾値４よりも大きいか否かを判断する（Ｓ２６４）。上記条件を満たす場合には（Ｓ２６４でＹＥＳ）、フレームＦＲＳからフレームＦＲＥまでを音楽調波構造性フレームであると判断する（Ｓ２６６）。

Ｓ２６０の条件を満たさない場合（Ｓ２６０でＮＯ）、またはＳ２６４でＮＯの場合、音楽調波構造は有しないが、調波構造を有するフレームであると判断できる。このため、フレームＦＲＳからフレームＦＲＥまでを非調波構造性フレームと判断し、カウンタＣＨに０を代入し、カウンタＣＮにＣＮ＋ＦＲＥ－ＦＲＳを代入する（Ｓ２６８）。

フレームワイズに調波性判断を行なう場合には調波構造性仮判定の判断を用い、より正確に調波性判断を行なう場合には調波構造性区間決定の結果を用いることにより、場合によりこれらを切り替えて使用するなどの自由度の高い選択が可能である。

上述したような処理を行なうことにより、調波構造性フレームと、音楽調波構造性フレームと、非調波構造性フレームと確定を行なうことができる。

以上説明したように、本実施の形態によると、入力される音声信号に対し、リアルタイムに調波構造性を有するか否かの判断を行なうことができる。このため、携帯電話などにおいて、所定フレーム遅れで非調波性のノイズを除去したりすることができる。また、音声と音楽とを見分けることができるため、携帯電話などを用いた通信において、音声部分と音楽部分とを異なる方法により符号化して通信を行なったりすることができる。

上述の実施の形態によると、環境雑音下で発声を行なった場合であっても、入力信号のレベル変動に依存せず、精度よく音声区間を決定する

ことができる。また、突発雑音や周期性雑音の影響を取り除き、精度良く音声区間を検出することができる。さらに、リアルタイムで音声区間を検出することができる。さらにまた、調波構造が小さい子音部分をも音声区間として精度良く検出することができる。また、入力信号を周波数変換したスペクトル成分にローカットフィルタをかけることにより、スペクトル包絡成分を除去することができる。

以上、本発明に係る音声区間検出装置について実施の形態１～５に基づいて説明したが、本発明はこれらの実施の形態に限定されるものではない。

（ＦＦＴ部２００の変形例）

たとえば、上述の実施の形態では、音響特徴量としてＦＦＴパワースペクトル成分を用いる方法について述べたが、ＦＦＴスペクトル成分そのものや、フレーム単位での自己相関関数や、時間軸上での線形予測残差のＦＦＴパワースペクトル成分を用いてもよい。また、ＦＦＴスペクトルからＦＦＴパワースペクトルを求める前に、各スペクトル成分を二乗するなどの方法により、極大値および極小値の差を拡大させ、調波構造を強調させてもよい。さらに、ＦＦＴスペクトルの対数を取り、ＦＦＴパワースペクトルを求める代わりに、ＦＦＴスペクトルの平方根を求め、ＦＦＴパワースペクトルとしてもよい。さらにまた、ＦＦＴスペクトル成分を求める前に、時間軸データに対して、フレームごとにハミング窓などの係数をかけてもよいし、プリエンファシス処理（１－ｚ－１）を行なうことで、高域強調を行ってもよい。また、音響特徴量として線スペクトル周波数（ＬＳＦ）を用いてもよい。また、周波数変換演算として、ＦＦＴに限られるものではなく、ＤＦＴ(Discrete Fourier Transform)、ＤＣＴ(Discrete Cosine Transform)、ＤＳＴ(Discrete Sine Transform)を用いても良い。

（調波構造抽出部２０１の変形例）

また、調波構造抽出部２０１によるスペクトル成分Ｓ（ｆ）に含まれるフロア成分の除去処理（図３のＳ２６）の代わりに、スペクトル成分Ｓ（ｆ）にローカットフィルタを通過させるようにしてもよい。各フレームのスペクトル成分Ｓ(ｆ)を周波数軸方向に並べた波形とみなすと、スペクトル包絡成分は、調波構造に比べゆっくりした変動である。このため、スペクトル成分にローカットフィルタをかけることにより、スペクトル包絡成分を除去することができる。この手法は時間軸上でローカットフィルタを用いて低周波数成分を取り除くことに相当するが、帯域パワーやスペクトル包絡などの情報と調音構造とを同時に評価することができる点において、周波数軸上で処理する方法の方が好ましいといえる。ただし、このようなローカットフィルタを用いて算出されたスペクトル成分は、調音構造に起因する変動の他に、非周期雑音や電子音などの単一周波数を有する音声以外の音を含んでいる可能性がある。しかし、これらの音は、有声評価部２１０や音声区間決定部２０５の処理により除去される。

その他のフロア成分除去の方法としては、各スペクトル成分のうち、所定の基準値以下のスペクトル成分は利用しないようにする方法がある。基準値の算出方法としては、全フレームのスペクトル成分の平均値を基準値に用いる方法、一発声の持続時間よりも十分に長い時間(たとえば、５秒間）におけるスペクトル成分の平均値を基準値に用いる方法、スペクトル成分をいくつかの帯域に予め分割しておき、帯域ごとにスペクトル成分の平均値を求める基準値とする方法などがある。特に、静かな環境からうるさい環境へ変化するなどの環境の変動がある場合には、基準値として、全フレームのスペクトル成分の平均値を利用するよりも、現在検出しようとしているフレームを含む数秒程度の区間のスペクトル成

分の平均値を用いるのがよい。

（特徴量フレーム間相関値算出部２０３の変形例）

また、特徴量フレーム間相関値算出部２０３は、相関関数として、式（３）の代わりに、次式（２４）を用いて相関値Ｅ１（ｊ）を求めるようにしてもよい。ここで、式（２４）は、Ｐ（ｉ－１）およびＰ（ｉ）を１２８次元ベクトル空間中のベクトルとした場合の２つのベクトルＰ（ｉ－１）およびＰ（ｉ）がなす角の余弦を示している。 また、特徴量フレーム間相関値算出部２０３は、相関値Ｅ１（ｊ）の代わりにフレームｊと４フレーム離れたフレーム間相関値を特徴とさせて、次式（２５）および（２６）に従い相関値Ｅ２（ｊ）を求めるようにしてもよいし、８フレーム離れたフレーム間相関値を特徴として、次式（２７）および（２８）に従い相関値Ｅ３（ｊ）を求めるようにしてもよい。このように、離れたフレーム間で相関値を求めることにより、突発的な環境雑音に強い相関値を得ることができるという特徴がある。

さらに、次式（２９）～（３１）に従い、相関値Ｅ１（ｊ）、相関値Ｅ２（ｊ）、相関値Ｅ３（ｊ）の大小関係に応じた相関値Ｅ４（ｊ）を求めるようにしてもよいし、次式（３２）に従い相関値Ｅ１（ｊ）、相関値Ｅ２（ｊ）、相関値Ｅ３（ｊ）を加算した相関値Ｅ５（ｊ）を求めるようにしてもよいし、次式（３３）に従い、相関値Ｅ１（ｊ）、相関値Ｅ２（ｊ）、相関値Ｅ３（ｊ）のうちの最大値を相関値Ｅ６（ｊ）を求めるようにしてもよい。

$$\mathrm{xcorr}(P(i-1),P(i)) = \frac{P(i-1)\bullet P(i)}{\|P(i-1)\|P(i)\|} = \frac{p1(j-1)\times p1(j)+p2(j-1)\times p2(j)+\ldots+p128(j-1)\times p128(j)}{\sqrt{p1(j-1)^2+p2(j-1)^2+\ldots+p128(j-1)^2}\sqrt{p1(j)^2+p2(j)^2+\ldots+p128(j)^2}} \quad \cdots (24)$$

$$z2(i) = \max(\mathrm{xcorr}(P(i-4), P(i))) \quad \cdots (25)$$

$$E2(j) = \sum_{i=j-2}^{j} z2(i) \quad \cdots (26)$$

$$z3(i) = \max(\mathrm{xcorr}(P(i-8), P(i))) \quad \cdots (27)$$

$$E3(j) = \sum_{i=j-2}^{j} z3(i) \quad \cdots (28)$$

$$E4(j) = z1(j) \quad \cdots (29)$$

$$\text{if } (z3(j) > 0.5)\ E4(j) = E4(j) + z1(j)/z3(j) \quad \cdots (30)$$

$$\text{if } (z2(j) > 0.5)\ E4(j) = E4(j) + z1(j)/z2(j) \quad \cdots (31)$$

$$E5(j) = E1(j) + E2(j) + E3(j)$$

$$= \sum_{i=j-2}^{j} z1(i) + \sum_{i=j-2}^{j} z2(i) + \sum_{i=j-2}^{j} z3(i) \quad \cdots (32)$$

$$E6(j) = \max(E1(j), E2(j), E3(j))$$

$$= \max\left(\sum_{i=j-2}^{j} z1(i), \sum_{i=j-2}^{j} z2(i), \sum_{i=j-2}^{j} z3(i)\right) \quad \cdots (33)$$

なお、相関値は、上述のＥ１（ｊ）～Ｅ６（ｊ）の６つに限定されるわけではなく、これらの相関値を組み合わせて、新たな相関値を算出するようにしてもよい。たとえば、過去に推定された入力音響信号のＳＮＲから、ＳＮＲが小さい場合には、相関値Ｅ１（ｊ）を使用し、ＳＮＲが大きい場合には、相関値Ｅ２（ｊ）またはＥ３（ｊ）を使用するようにしてもよい。

（音声区間決定部２０５の変形例）

図６を用いて説明した音声区間決定部２０５の処理は、相関値による有声区間決定処理（Ｓ４２～Ｓ５０）、有声区間の連結処理（Ｓ５２～Ｓ５８）、および有声区間の持続時間による音声区間決定処理（Ｓ６０～Ｓ６８）の３つの処理に大きく分類されるが、これら３つの処理を図６に示される順序で実行する必要はなく、他の順序で実行するようにしてもよい。また、３つの処理のうち、１つまたは２つの処理のみを実行するようにしてもよい。また、図６は、一発声単位で処理を行なう例であるが、たとえば注目フレームごとに相関値による有声区間決定処理のみを行なうことで、フレーム単位で音声区間を決定補正してもよい。さらに、リアルタイム性が要求されることを想定して、フレーム単位の相関値による音声区間を速報値として出力しておき、別途、定期的に、一発声等長い単位で補正決定された音声区間を確定値として出力することで、リアルタイム性にも、検出区間性能にも対応可能な、音声検出器として作用させてもよい。

（ＳＮＲ推定部２０６の変形例）

また、ＳＮＲ推定部２０６は、入力信号から直接ＳＮＲを推定するようにしてもよい。たとえば、差分処理部２０４で算出された補正相関値が正の部分をＳ（シグナル）部分とし、Ｓ部分のパワーを求め、補正相関値が負の部分をＮ（ノイズ）部分とし、Ｎ部分のパワーを求め、ＳＮＲを求めるようにする。

（その他の変形例）

さらに、上述の音声区間検出処理を前処理とし、音声区間のみについて音声認識を行なう音声認識装置に音声区間検出装置を使用してもよい。

また、上述の音声区間検出処理を前処理として、音声区間のみについて録音を行なうＩＣ（Integrated Circuit）レコーダなどの音声録音装置に音声区間検出装置を使用しても良い。このように、音声区間のみを

録音することにより、ＩＣレコーダの記憶領域を効率的に利用することが可能となる。再生時には、音声区間のみを抽出し、話速変換機能を用いて、効率的な再生も可能となる。

また、音声区間以外の区間の入力信号をカットして雑音を抑制する雑音抑制装置に音声認識装置を利用してもよい。

さらにまた、ＶＴＲ（Video Tape Recorder）等で撮影された映像から、音声区間の映像を抽出するのに、上述の音声区間検出処理を用いてもよく、映像を編集するオーサリングツールなどにも適用可能である。

また、図４（ｆ）に示されるパワースペクトル成分Ｓ’（ｆ）のうち、調波構造が最もよく保たれている帯域を１つ以上抽出し、その帯域のみを用いて処理を行なうようにしてもよい。

また、非音声区間を検出することにより、非音声区間内でノイズの特徴を学習し、ノイズ除去のためのフィルタリング係数、ノイズ決定のパラメータ等を決めたりするようにしてもよい。このようにすることにより、ノイズ除去のための装置を作成することができる。

また、上述した実施の形態における各種調波構造性値または各種相関値と、各種音声区間決定方法との組み合わせは、上述した実施の形態に限定されない。

産業上の利用の可能性

本発明に係る音声区間検出装置は、音声区間と雑音区間との精度よい選別が可能となるため、音声認識装置の前処理装置、音声区間のみを録音するＩＣレコーダ、音声区間と音楽区間とを異なる符号化方法で符号化する通信装置等に有用である。

請　求　の　範　囲

1．　入力音響信号から音声が含まれる区間を音声区間として検出する調波構造性音響信号区間検出方法であって、

前記入力音響信号に対し、所定の時間で区切られたフレーム単位で音響特徴量を抽出する音響特徴量抽出ステップと、

前記音響特徴量の持続性を評価し、評価結果に従って音声区間を決定する区間決定ステップとを含み、

前記音響特徴量抽出ステップでは、所定の時間で区切られたフレーム単位で前記入力音響信号を周波数変換し、調波構造を尺度化した音響特徴量を抽出し、

前記区間決定ステップでは、前記音響特徴量の同一フレーム内における相関値または前記音響特徴量の異なるフレーム間における相関値に基づいて、音声区間を決定する

ことを特徴とする調波構造性音響信号区間検出方法。

2．　前記音響特徴量抽出ステップでは、さらに、前記周波数変換の結果より調波構造を強調し、前記音響特徴量を抽出する

ことを特徴とする請求の範囲第１項に記載の調波構造性音響信号区間検出方法。

3．　前記音響特徴量抽出ステップでは、さらに、前記周波数変換の結果より調波構造を抽出し、当該調波構造を含む所定帯域の周波数変換の結果を、前記音響特徴量とする

ことを特徴とする請求の範囲第２項に記載の調波構造性音響信号区間検出方法。

4. 前記音響特徴量抽出ステップでは、さらに、フレーム単位の周波数変換の結果を所定の周波数帯域幅ごとに分割し、同一フレーム内の所定の周波数帯域間で、前記周波数変換の結果の相関値を算出し、算出結果に基づいて前記音響特徴量を抽出する

ことを特徴とする請求の範囲第1項に記載の調波構造性音響信号区間検出方法。

5. 前記音響特徴量抽出ステップでは、さらに、フレームごとに前記相関値の最大値と最小値との差を求め、当該差に基づいて、前記音響特徴量を抽出する

ことを特徴とする請求の範囲第4項に記載の調波構造性音響信号区間検出方法。

6. 前記音響特徴量抽出ステップでは、フレーム単位の周波数変換の結果を所定の周波数帯域幅ごとに分割し、異なるフレーム間、かつ所定の周波数帯域間で、前記周波数変換の結果の相関値を算出し、算出結果に基づいて前記音響特徴量を抽出する

ことを特徴とする請求の範囲第1項に記載の調波構造性音響信号区間検出方法。

7. 前記音響特徴量抽出ステップでは、さらに、フレームごとに前記相関値の最大値と最小値との差を求め、前記差に基づいて、前記音響特徴量を抽出する

ことを特徴とする請求の範囲第6項に記載の調波構造性音響信号区間検出方法。

8. 前記区間決定ステップでは、異なるフレーム間における前記音響特徴量の相関値に基づいて、前記音響特徴量の持続性を評価し、評価結果に従って音声区間を決定する

ことを特徴とする請求の範囲第１項に記載の調波構造性音響信号区間検出方法。

9. 前記区間決定ステップでは、異なるフレーム間における前記音響特徴量の分散に基づいて、前記音響特徴量の持続性を評価し、評価結果に従って音声区間を決定する

ことを特徴とする請求の範囲第１項に記載の調波構造性音響信号区間検出方法。

１０. 前記音響特徴量の持続性を評価する評価値を算出する評価ステップと、

前記評価値の時間的な連続性を評価し、評価結果に従って音声区間を決定する音声区間決定ステップとを含む

ことを特徴とする請求の範囲第１項に記載の調波構造性音響信号区間検出方法。

１１. 前記区間決定ステップは、さらに、

所定数のフレームにわたる、音響特徴量抽出ステップにおいて算出される音響特徴量または、前記評価ステップにおいて算出される前記評価値と、第１の所定しきい値との比較に基づいて、前記入力音響信号の音声雑音比を推定するステップと、

推定された前記音声雑音比が第２の所定しきい値以上の場合には、前記評価ステップにおいて算出される前記評価値に基づいて前記音声区間

を決定するステップとを含み、

前記音声区間決定ステップでは、前記音声雑音比が前記第２の所定しきい値未満の場合に、前記評価値の時間的な連続性を評価し、評価結果に従って前記音声区間を決定する

ことを特徴とする請求の範囲第１０項に記載の調波構造性音響信号区間検出方法。

１２．　前記区間決定ステップは、

前記音響特徴量の持続性を評価する評価値を算出する評価ステップと、

前記評価値の時間的な連続性を評価し、評価結果に従って調波構造を有するが音声ではない非音声調波構造区間を決定する非音声調波構造区間決定ステップとを含む

ことを特徴とする請求の範囲第１項に記載の調波構造音響調波構造性音響信号区間検出方法。

１３．　前記音響特徴量抽出ステップは、

所定の時間で区切られたフレーム単位で、前記入力音響信号を周波数変換する周波数変換ステップと、

フレーム単位の周波数変換の結果を所定の周波数帯域ごとに分割し、同一フレーム内の所定の周波数帯域間で、前記周波数変換の結果の相関値を算出する相関値算出ステップと、

同一フレーム内における相関値の最大値または最小値をとる周波数帯域の識別子を前記音響特徴量として抽出する抽出ステップとを含む

ことを特徴とする請求の範囲第１２項に記載の調波構造性音響信号区間検出方法。

１４．　前記音響特徴量抽出ステップは、

所定の時間で区切られたフレーム単位で、前記入力音響信号を周波数変換する周波数変換ステップと、

所定数離れたフレーム間で前記周波数変換の結果の相関値を算出する相関値算出ステップと、

所定数のフレームごとに前記相関値の分散を算出することにより、前記調波構造を尺度化した音響特徴量を抽出する音響特徴量抽出ステップとを含む

ことを特徴とする請求の範囲第１項に記載の調波構造音響区間検出方法。

１５．　前記区間決定ステップでは、２種以上の異なる間隔のフレーム間における複数の相関値に基づいて、前記持続性を評価する

ことを特徴とする請求の範囲第１項に記載の調波構造性音響信号区間検出方法。

１６．　前記区間決定ステップでは、前記入力音響信号の音声雑音比に基づいて、前記２種以上の異なる間隔のフレーム間における相関値のいずれかを選択し、選択された相関値に基づいて前記持続性を評価する

ことを特徴とする請求の範囲第１５項に記載の調波構造性音響信号区間検出方法。

１７．　前記区間決定ステップでは、前記音響特徴量のフレーム間における相関値と、前記相関値を所定フレーム数にわたり平均した平均値との補正相関値に基づいて、前記持続性を評価する

ことを特徴とする請求の範囲第１項に記載の調波構造性音響信号区間

検出方法。

１８．　前記音響特徴量抽出手段は、所定の時間で区切られたフレーム単位で前記入力音響信号を周波数変換し、調波構造を尺度化した音響特徴量を抽出し、

前記区間決定手段は、前記音響特徴量の同一フレーム内における相関値または前記音響特徴量の異なるフレーム間における相関値に基づいて、音声区間を決定する

ことを特徴とする請求の範囲第２８項に記載の調波構造性音響信号区間検出装置。

１９．　入力音響信号に含まれる音声を認識する音声認識装置であって、

前記入力音響信号に対し、所定の時間で区切られたフレーム単位で音響特徴量を抽出する音響特徴量抽出手段と、

前記音響特徴量の持続性を評価し、評価結果に従って音声区間を決定する区間決定手段と、

前記区間決定手段で決定された音声区間において音声認識を行なう認識手段とを備え、

前記音響特徴量抽出手段は、所定の時間で区切られたフレーム単位で前記入力音響信号を周波数変換し、調波構造を尺度化した音響特徴量を抽出し、

前記区間決定手段は、前記音響特徴量の同一フレーム内における相関値または前記音響特徴量の異なるフレーム間における相関値に基づいて、音声区間を決定する

ことを特徴とする音声認識装置。

２０．　入力音響信号に含まれる音声を録音する音声録音装置であって、

前記入力音響信号に対し、所定の時間で区切られたフレーム単位で音響特徴量を抽出する音響特徴量抽出手段と、

前記音響特徴量の持続性を評価し、評価結果に従って音声区間を決定する区間決定手段と、

前記区間決定手段で決定された音声区間における入力音響信号を録音する録音手段とを備え、

前記音響特徴量抽出手段は、所定の時間で区切られたフレーム単位で前記入力音響信号を周波数変換し、調波構造を尺度化した音響特徴量を抽出し、

前記区間決定手段は、前記音響特徴量の同一フレーム内における相関値または前記音響特徴量の異なるフレーム間における相関値に基づいて、音声区間を決定する

ことを特徴とする音声録音装置。

２１．　入力音響信号に対し、所定の時間で区切られたフレーム単位で音響特徴量を抽出する音響特徴量抽出ステップと、

前記音響特徴量の持続性を評価し、評価結果に従って音声区間を決定する区間決定ステップとをコンピュータに実行させ、

前記音響特徴量抽出ステップでは、所定の時間で区切られたフレーム単位で前記入力音響信号を周波数変換し、調波構造を尺度化した音響特徴量を抽出し、

前記区間決定ステップでは、前記音響特徴量の同一フレーム内における相関値または前記音響特徴量の異なるフレーム間における相関値に基づいて、音声区間を決定する

ことを特徴とするプログラム。

補正書の請求の範囲

［２００４年１１月１日（０１．１１．０４）国際事務局受理：出願当初の請求の範囲１８は補正された；他の請求の範囲は変更なし。（８頁）］

１．　入力音響信号から音声が含まれる区間を音声区間として検出する調波構造性音響信号区間検出方法であって、

前記入力音響信号に対し、所定の時間で区切られたフレーム単位で音響特徴量を抽出する音響特徴量抽出ステップと、

前記音響特徴量の持続性を評価し、評価結果に従って音声区間を決定する区間決定ステップとを含み、

前記音響特徴量抽出ステップでは、所定の時間で区切られたフレーム単位で前記入力音響信号を周波数変換し、調波構造を尺度化した音響特徴量を抽出し、

前記区間決定ステップでは、前記音響特徴量の同一フレーム内における相関値または前記音響特徴量の異なるフレーム間における相関値に基づいて、音声区間を決定する

ことを特徴とする調波構造性音響信号区間検出方法。

２．　前記音響特徴量抽出ステップでは、さらに、前記周波数変換の結果より調波構造を強調し、前記音響特徴量を抽出する

ことを特徴とする請求の範囲第１項に記載の調波構造性音響信号区間検出方法。

３．　前記音響特徴量抽出ステップでは、さらに、前記周波数変換の結果より調波構造を抽出し、当該調波構造を含む所定帯域の周波数変換の結果を、前記音響特徴量とする

ことを特徴とする請求の範囲第２項に記載の調波構造性音響信号区間検出方法。

4. 前記音響特徴量抽出ステップでは、さらに、フレーム単位の周波数変換の結果を所定の周波数帯域幅ごとに分割し、同一フレーム内の所定の周波数帯域間で、前記周波数変換の結果の相関値を算出し、算出結果に基づいて前記音響特徴量を抽出する

ことを特徴とする請求の範囲第1項に記載の調波構造性音響信号区間検出方法。

5. 前記音響特徴量抽出ステップでは、さらに、フレームごとに前記相関値の最大値と最小値との差を求め、当該差に基づいて、前記音響特徴量を抽出する

ことを特徴とする請求の範囲第4項に記載の調波構造性音響信号区間検出方法。

6. 前記音響特徴量抽出ステップでは、フレーム単位の周波数変換の結果を所定の周波数帯域幅ごとに分割し、異なるフレーム間、かつ所定の周波数帯域間で、前記周波数変換の結果の相関値を算出し、算出結果に基づいて前記音響特徴量を抽出する

ことを特徴とする請求の範囲第1項に記載の調波構造性音響信号区間検出方法。

7. 前記音響特徴量抽出ステップでは、さらに、フレームごとに前記相関値の最大値と最小値との差を求め、前記差に基づいて、前記音響特徴量を抽出する

ことを特徴とする請求の範囲第6項に記載の調波構造性音響信号区間検出方法。

８．　前記区間決定ステップでは、異なるフレーム間における前記音響特徴量の相関値に基づいて、前記音響特徴量の持続性を評価し、評価結果に従って音声区間を決定する

ことを特徴とする請求の範囲第１項に記載の調波構造性音響信号区間検出方法。

９．　前記区間決定ステップでは、異なるフレーム間における前記音響特徴量の分散に基づいて、前記音響特徴量の持続性を評価し、評価結果に従って音声区間を決定する

ことを特徴とする請求の範囲第１項に記載の調波構造性音響信号区間検出方法。

１０．　前記音響特徴量の持続性を評価する評価値を算出する評価ステップと、

前記評価値の時間的な連続性を評価し、評価結果に従って音声区間を決定する音声区間決定ステップとを含む

ことを特徴とする請求の範囲第１項に記載の調波構造性音響信号区間検出方法。

１１．　前記区間決定ステップは、さらに、

所定数のフレームにわたる、音響特徴量抽出ステップにおいて算出される音響特徴量または、前記評価ステップにおいて算出される前記評価値と、第１の所定しきい値との比較に基づいて、前記入力音響信号の音声雑音比を推定するステップと、

推定された前記音声雑音比が第２の所定しきい値以上の場合には、前記評価ステップにおいて算出される前記評価値に基づいて前記音声区間

を決定するステップとを含み、

前記音声区間決定ステップでは、前記音声雑音比が前記第２の所定しきい値未満の場合に、前記評価値の時間的な連続性を評価し、評価結果に従って前記音声区間を決定する

ことを特徴とする請求の範囲第１０項に記載の調波構造性音響信号区間検出方法。

１２．　前記区間決定ステップは、

前記音響特徴量の持続性を評価する評価値を算出する評価ステップと、

前記評価値の時間的な連続性を評価し、評価結果に従って調波構造を有するが音声ではない非音声調波構造区間を決定する非音声調波構造区間決定ステップとを含む

ことを特徴とする請求の範囲第１項に記載の調波構造音響調波構造性音響信号区間検出方法。

１３．　前記音響特徴量抽出ステップは、

所定の時間で区切られたフレーム単位で、前記入力音響信号を周波数変換する周波数変換ステップと、

フレーム単位の周波数変換の結果を所定の周波数帯域ごとに分割し、同一フレーム内の所定の周波数帯域間で、前記周波数変換の結果の相関値を算出する相関値算出ステップと、

同一フレーム内における相関値の最大値または最小値をとる周波数帯域の識別子を前記音響特徴量として抽出する抽出ステップとを含む

ことを特徴とする請求の範囲第１２項に記載の調波構造性音響信号区間検出方法。

１４．　前記音響特徴量抽出ステップは、

所定の時間で区切られたフレーム単位で、前記入力音響信号を周波数変換する周波数変換ステップと、

所定数離れたフレーム間で前記周波数変換の結果の相関値を算出する相関値算出ステップと、

所定数のフレームごとに前記相関値の分散を算出することにより、前記調波構造を尺度化した音響特徴量を抽出する音響特徴量抽出ステップとを含む

ことを特徴とする請求の範囲第１項に記載の調波構造音響区間検出方法。

１５．　前記区間決定ステップでは、２種以上の異なる間隔のフレーム間における複数の相関値に基づいて、前記持続性を評価する

ことを特徴とする請求の範囲第１項に記載の調波構造性音響信号区間検出方法。

１６．　前記区間決定ステップでは、前記入力音響信号の音声雑音比に基づいて、前記２種以上の異なる間隔のフレーム間における相関値のいずれかを選択し、選択された相関値に基づいて前記持続性を評価する

ことを特徴とする請求の範囲第１５項に記載の調波構造性音響信号区間検出方法。

１７．　前記区間決定ステップでは、前記音響特徴量のフレーム間における相関値と、前記相関値を所定フレーム数にわたり平均した平均値との補正相関値に基づいて、前記持続性を評価する

ことを特徴とする請求の範囲第１項に記載の調波構造性音響信号区間

検出方法。

１８．　（補正後）入力音響信号から音声が含まれる区間を音声区間として検出する調波構造性音響信号区間検出装置であって、

前記入力音響信号に対し、所定の時間で区切られたフレーム単位で音響特徴量を抽出する音響特徴量抽出手段と、

前記音響特徴量の持続性を評価し、評価結果に従って音声区間を決定する区間決定手段とを含み、

前記音響特徴量抽出手段は、所定の時間で区切られたフレーム単位で前記入力音響信号を周波数変換し、調波構造を尺度化した音響特徴量を抽出し、

前記区間決定手段は、前記音響特徴量の同一フレーム内における相関値または前記音響特徴量の異なるフレーム間における相関値に基づいて、音声区間を決定する

ことを特徴とする調波構造性音響信号区間検出装置。

１９．　入力音響信号に含まれる音声を認識する音声認識装置であって、

前記入力音響信号に対し、所定の時間で区切られたフレーム単位で音響特徴量を抽出する音響特徴量抽出手段と、

前記音響特徴量の持続性を評価し、評価結果に従って音声区間を決定する区間決定手段と、

前記区間決定手段で決定された音声区間において音声認識を行なう認識手段とを備え、

前記音響特徴量抽出手段は、所定の時間で区切られたフレーム単位で前記入力音響信号を周波数変換し、調波構造を尺度化した音響特徴量を

抽出し、

前記区間決定手段は、前記音響特徴量の同一フレーム内における相関値または前記音響特徴量の異なるフレーム間における相関値に基づいて、音声区間を決定する

ことを特徴とする音声認識装置。

２０．　入力音響信号に含まれる音声を録音する音声録音装置であって、

前記入力音響信号に対し、所定の時間で区切られたフレーム単位で音響特徴量を抽出する音響特徴量抽出手段と、

前記音響特徴量の持続性を評価し、評価結果に従って音声区間を決定する区間決定手段と、

前記区間決定手段で決定された音声区間における入力音響信号を録音する録音手段とを備え、

前記音響特徴量抽出手段は、所定の時間で区切られたフレーム単位で前記入力音響信号を周波数変換し、調波構造を尺度化した音響特徴量を抽出し、

前記区間決定手段は、前記音響特徴量の同一フレーム内における相関値または前記音響特徴量の異なるフレーム間における相関値に基づいて、音声区間を決定する

ことを特徴とする音声録音装置。

２１．　入力音響信号に対し、所定の時間で区切られたフレーム単位で音響特徴量を抽出する音響特徴量抽出ステップと、

前記音響特徴量の持続性を評価し、評価結果に従って音声区間を決定する区間決定ステップとをコンピュータに実行させ、

前記音響特徴量抽出ステップでは、所定の時間で区切られたフレーム単位で前記入力音響信号を周波数変換し、調波構造を尺度化した音響特徴量を抽出し、

前記区間決定ステップでは、前記音響特徴量の同一フレーム内における相関値または前記音響特徴量の異なるフレーム間における相関値に基づいて、音声区間を決定する

ことを特徴とするプログラム。

条約１９条に基づく説明書

補正後の請求の範囲第１８項は、補正前の請求の範囲第１項に記載の調波構造性音響信号区間検出方法に対応する調波構造性音響信号区間検出装置の請求の範囲である。

補正後の請求の範囲第１８項は、補正前の請求の範囲第１項に記載の調波構造性音響信号区間検出方法に含まれる「ステップ」の文言を、「手段」に置き換えたものである。したがって、請求の範囲第１８項に対する補正は、新規事項の追加ではない。

図1

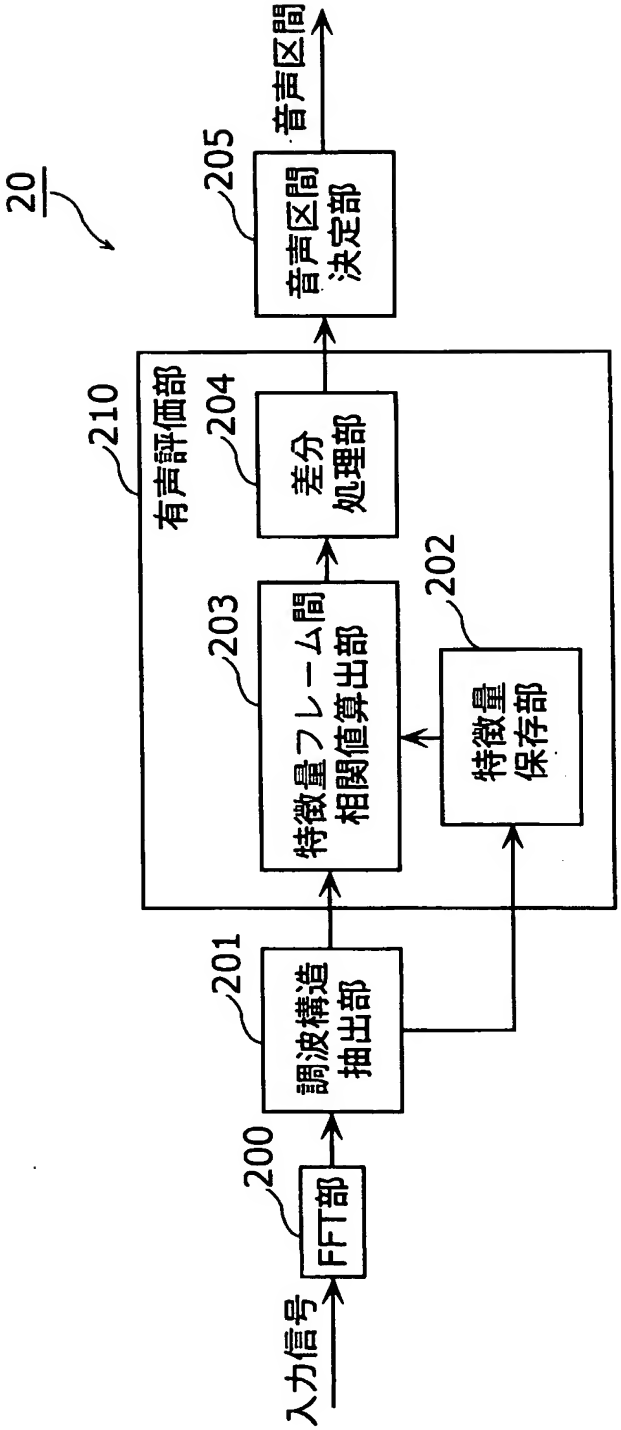

20
入力信号
FFT部
200
調波構造抽出部
201
特徴量フレーム間相関値算出部
203
特徴量保存部
202
差分処理部
204
有声評価部
210
音声区間決定部
205
音声区間

図2

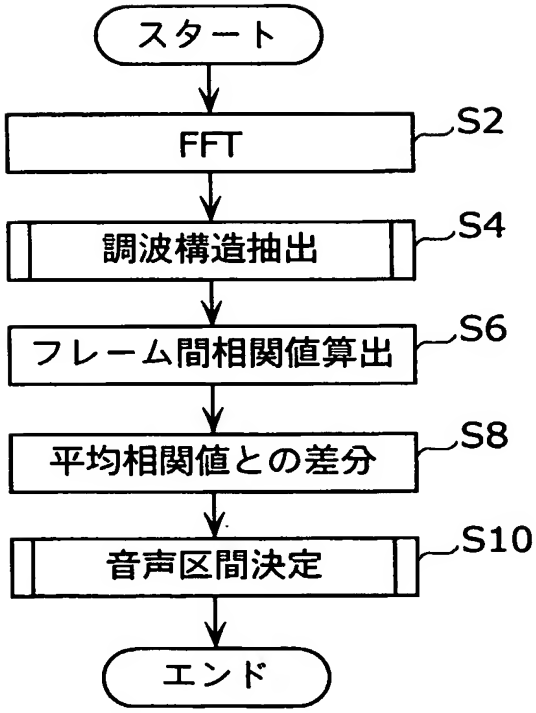

スタート
FFT
S2
調波構造抽出
S4
フレーム間相関値算出
S6
平均相関値との差分
S8
音声区間決定
S10
エンド

図3

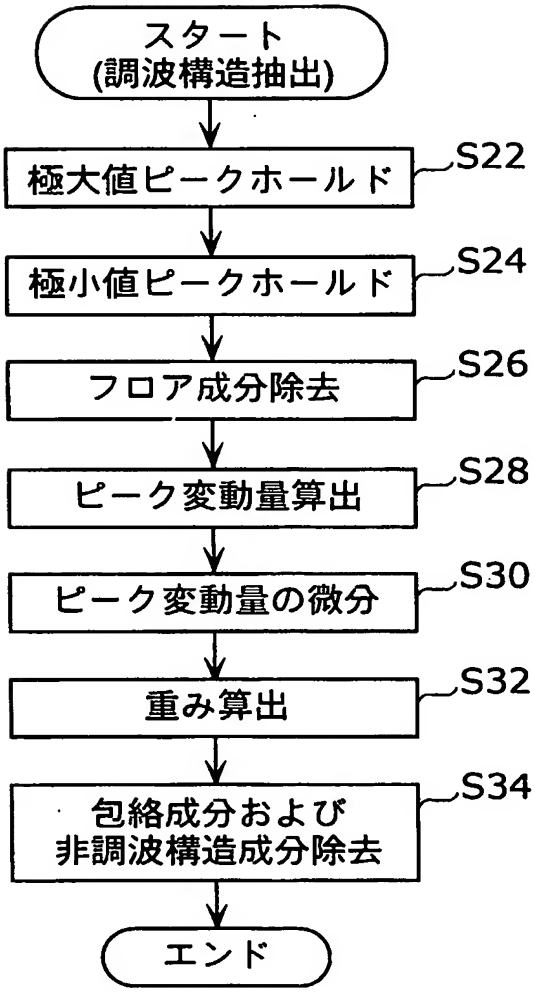

スタート
(調波構造抽出)
極大値ピークホールド
S22
極小値ピークホールド
S24
フロア成分除去
S26
ピーク変動量算出
S28
ピーク変動量の微分
S30
重み算出
S32
包絡成分および
非調波構造成分除去
S34
エンド

図4

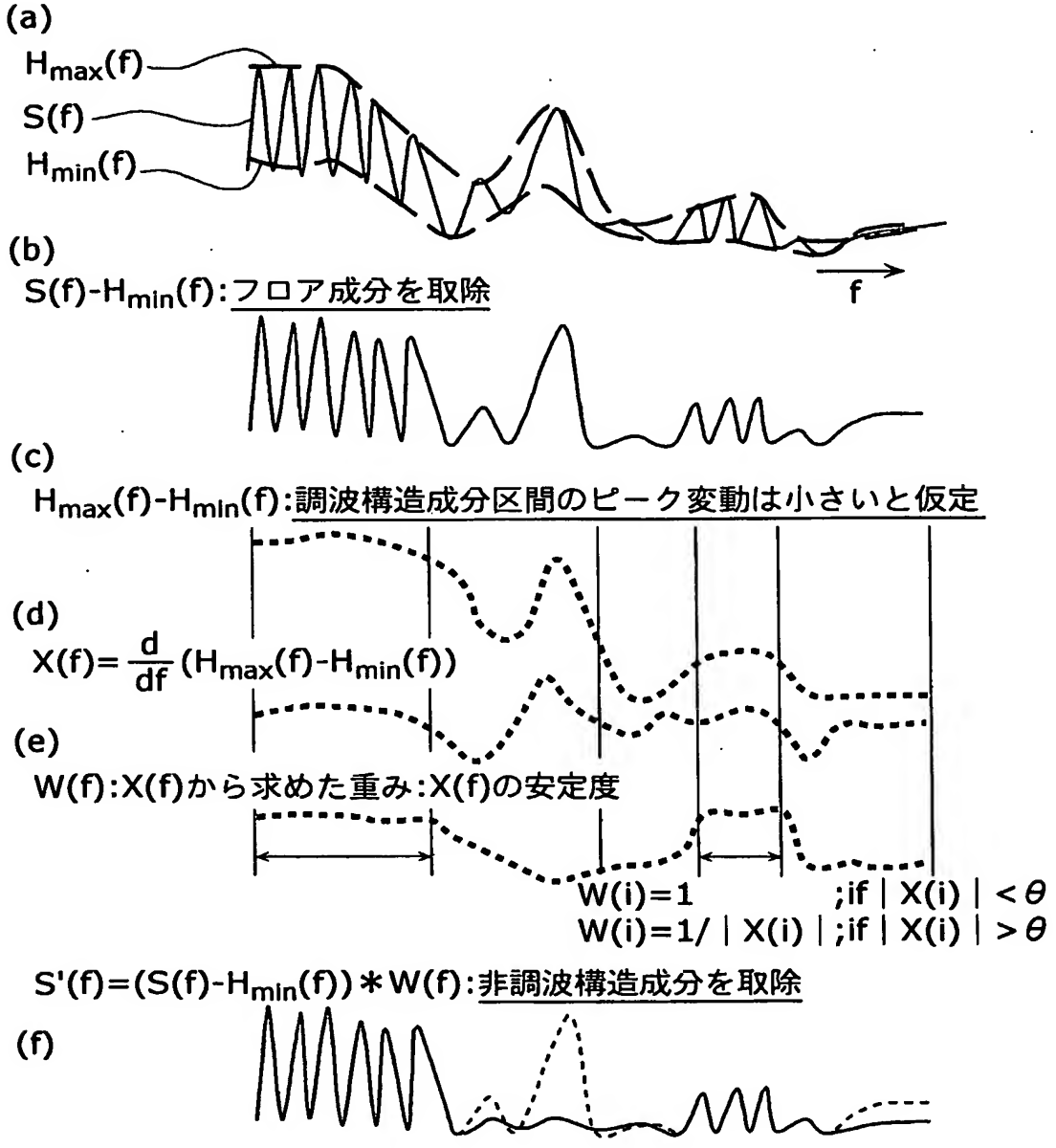

(a)
$H_{max}(f)$
S(f)
$H_{min}(f)$
(b)
$S(f)-H_{min}(f)$:フロア成分を取除
f
(c)
$H_{max}(f)-H_{min}(f)$:調波構造成分区間のピーク変動は小さいと仮定
(d)
$X(f)=\frac{d}{df}(H_{max}(f)-H_{min}(f))$
(e)
W(f):X(f)から求めた重み:X(f)の安定度
W(i)=1 ;if | X(i) | < θ
W(i)=1/ | X(i) | ;if | X(i) | > θ
$S'(f)=(S(f)-H_{min}(f))*W(f)$:非調波構造成分を取除
(f)

図5

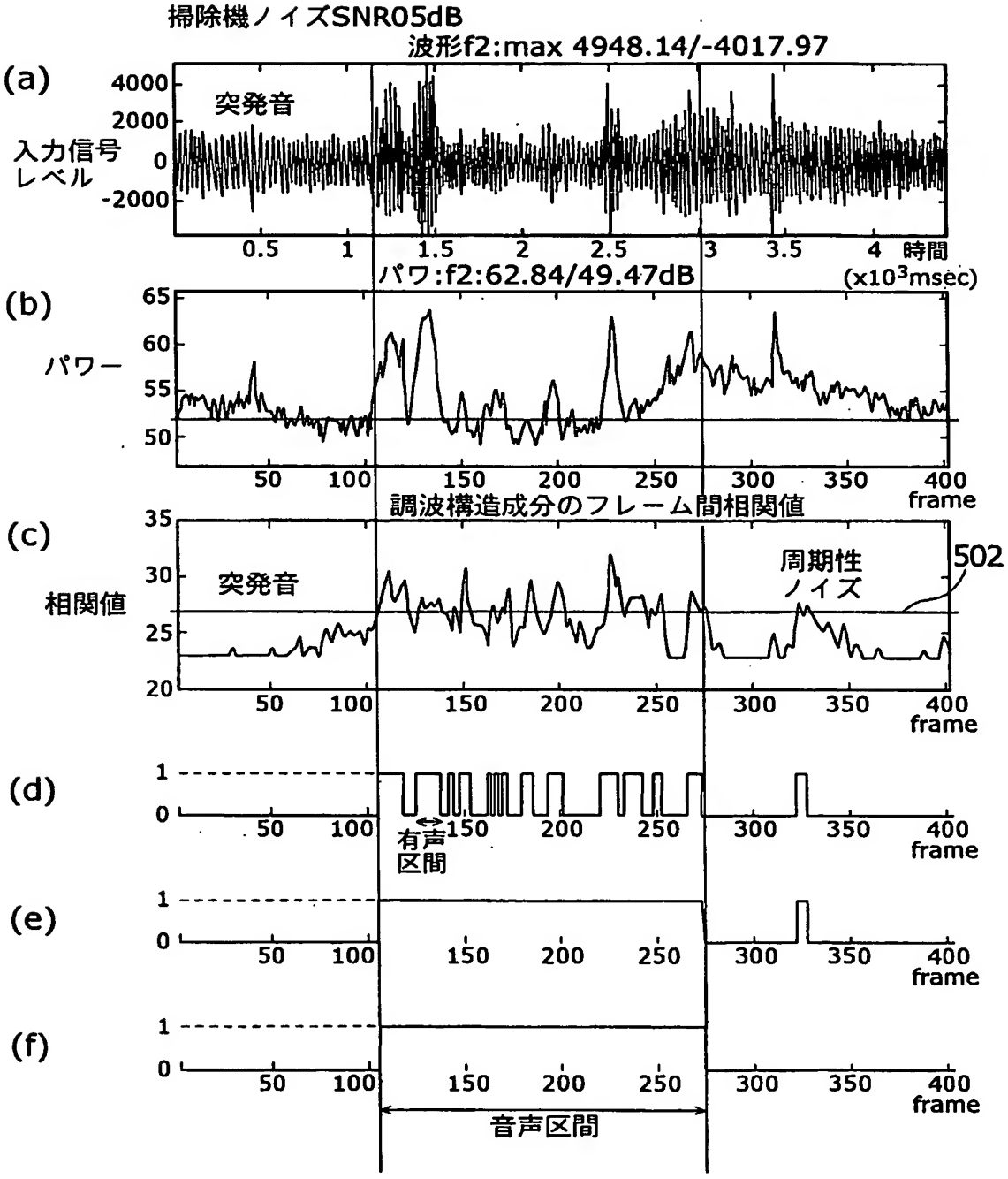

掃除機ノイズSNR05dB
波形f2:max 4948.14/-4017.97
(a)
入力信号レベル
突発音
4000
2000
0
-2000
0.5
1
1.5
2
2.5
3
3.5
4
時間
パワ:f2:62.84/49.47dB
(x10³msec)
(b)
パワー
65
60
55
50
50
100
150
200
250
300
350
400
frame
調波構造成分のフレーム間相関値
(c)
相関値
突発音
周期性ノイズ
502
35
30
25
20
(d)
1
0
有声区間
(e)
(f)
音声区間

図6

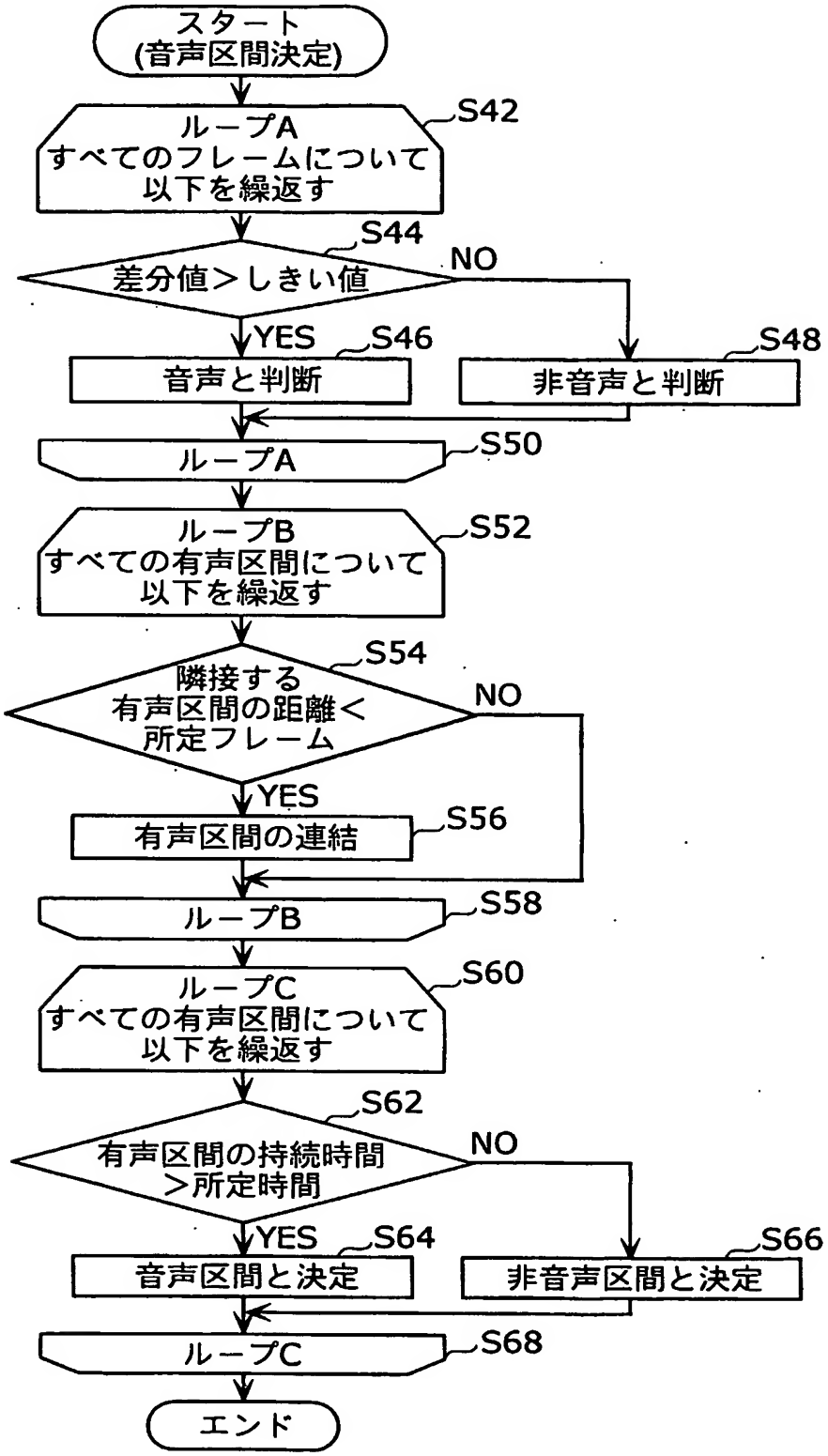

スタート
(音声区間決定)
ループA
すべてのフレームについて
以下を繰返す
S42
S44
差分値＞しきい値
NO
YES
S46
音声と判断
S48
非音声と判断
ループA
S50
ループB
すべての有声区間について
以下を繰返す
S52
S54
隣接する
有声区間の距離＜
所定フレーム
NO
YES
有声区間の連結
S56
ループB
S58
ループC
すべての有声区間について
以下を繰返す
S60
S62
有声区間の持続時間
＞所定時間
NO
YES
S64
音声区間と決定
S66
非音声区間と決定
ループC
S68
エンド

図7

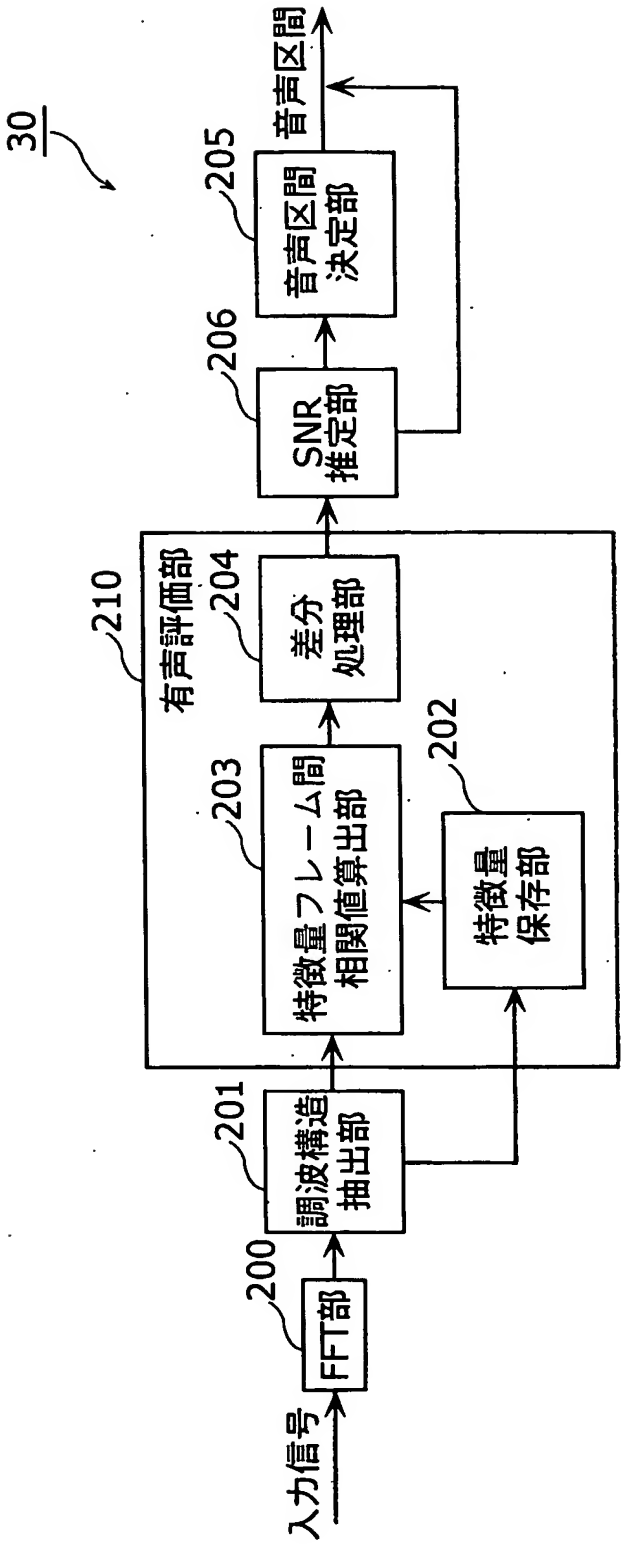

30
音声区間
205
音声区間決定部
206
SNR推定部
210
有声評価部
204
差分処理部
203
特徴量フレーム間相関値算出部
202
特徴量保存部
201
調波構造抽出部
200
FFT部
入力信号

図8

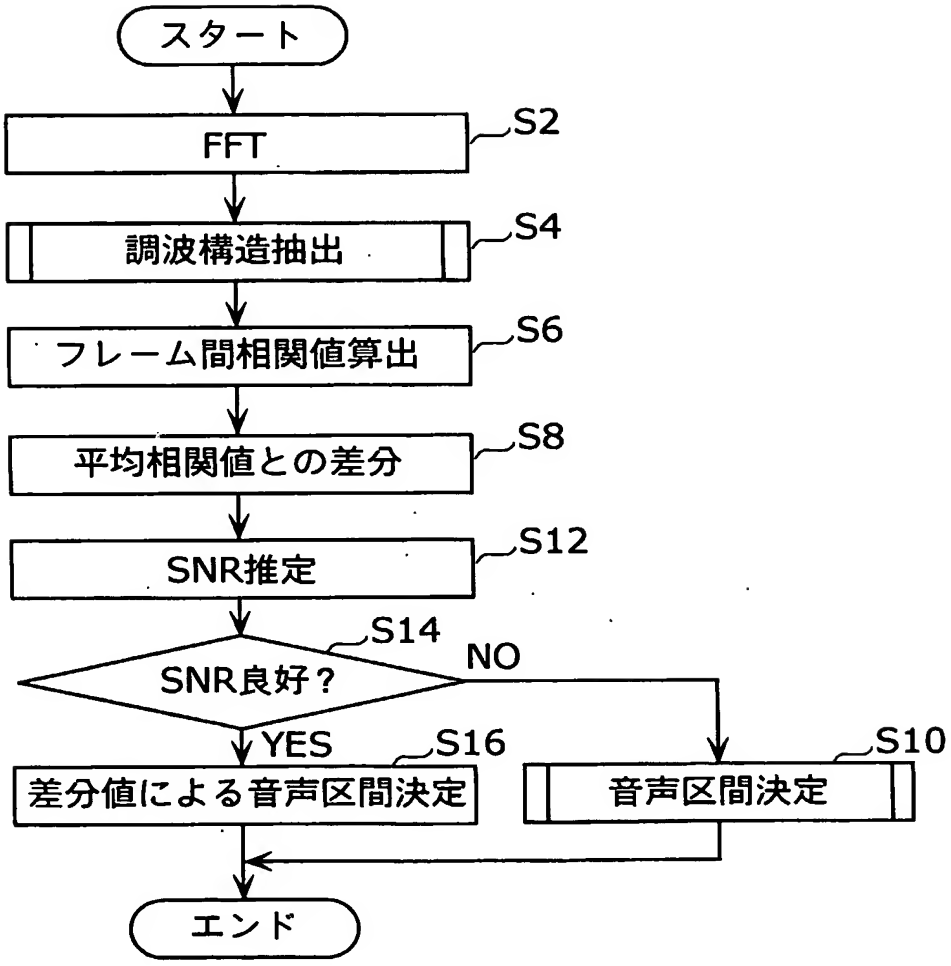

スタート
FFT
S2
調波構造抽出
S4
フレーム間相関値算出
S6
平均相関値との差分
S8
SNR推定
S12
S14
SNR良好？
NO
YES
S16
差分値による音声区間決定
S10
音声区間決定
エンド

図9

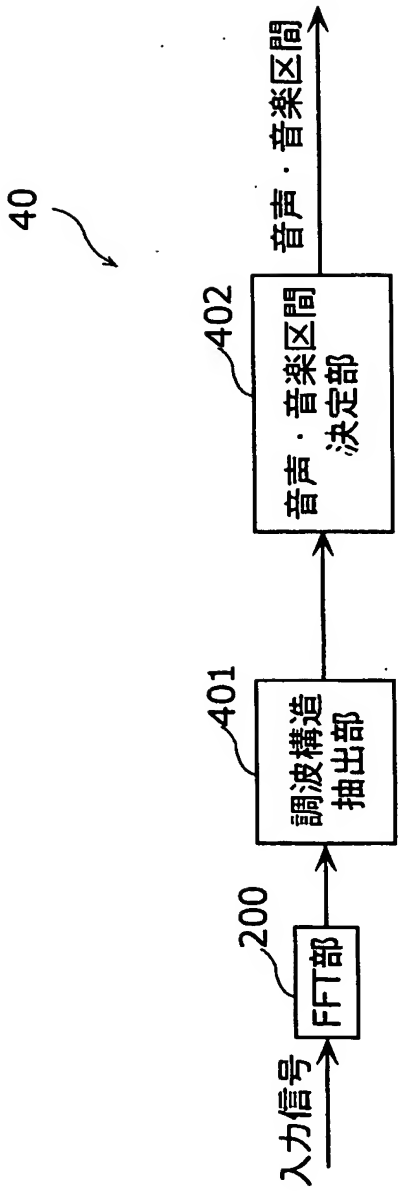


図10

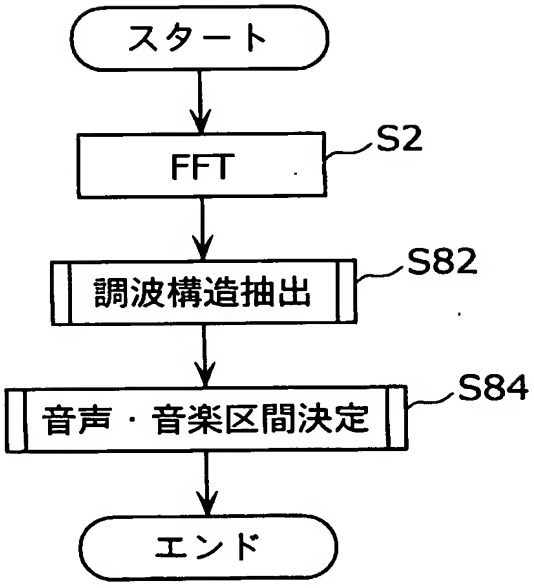


図11

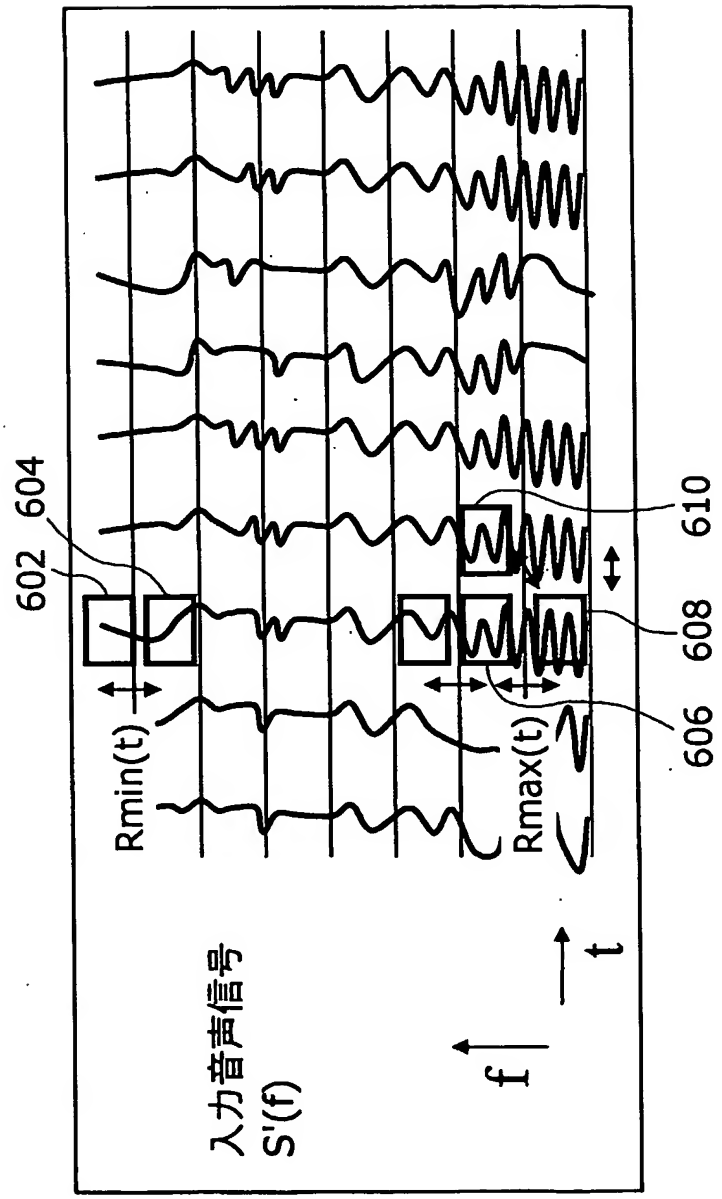


図12

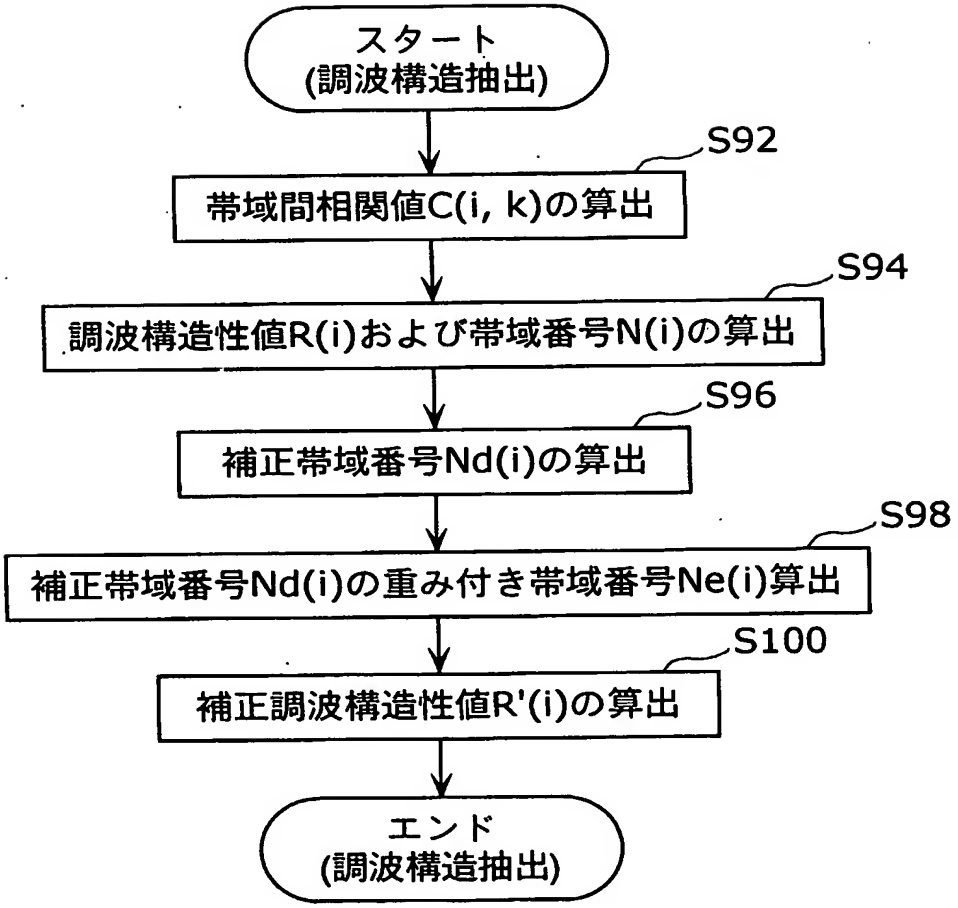

スタート
(調波構造抽出)
S92
帯域間相関値C(i, k)の算出
S94
調波構造性値R(i)および帯域番号N(i)の算出
S96
補正帯域番号Nd(i)の算出
S98
補正帯域番号Nd(i)の重み付き帯域番号Ne(i)算出
S100
補正調波構造性値R'(i)の算出
エンド
(調波構造抽出)

図13

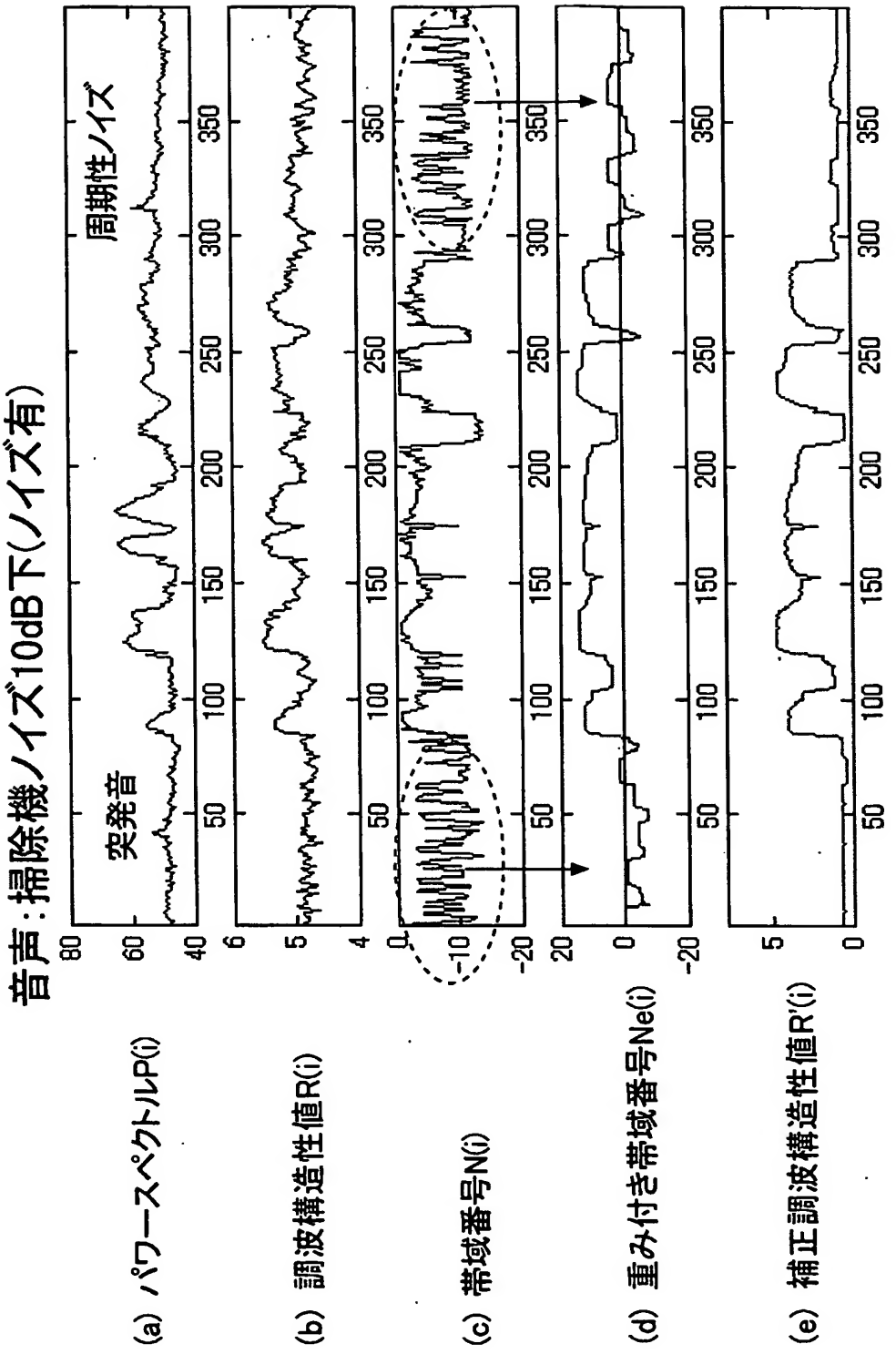

音声:掃除機ノイズ10dB下(ノイズ有)
突発音
周期性ノイズ
(a) パワースペクトルP(i)
(b) 調波構造性値R(i)
(c) 帯域番号N(i)
(d) 重み付き帯域番号Ne(i)
(e) 補正調波構造性値R'(i)

図14

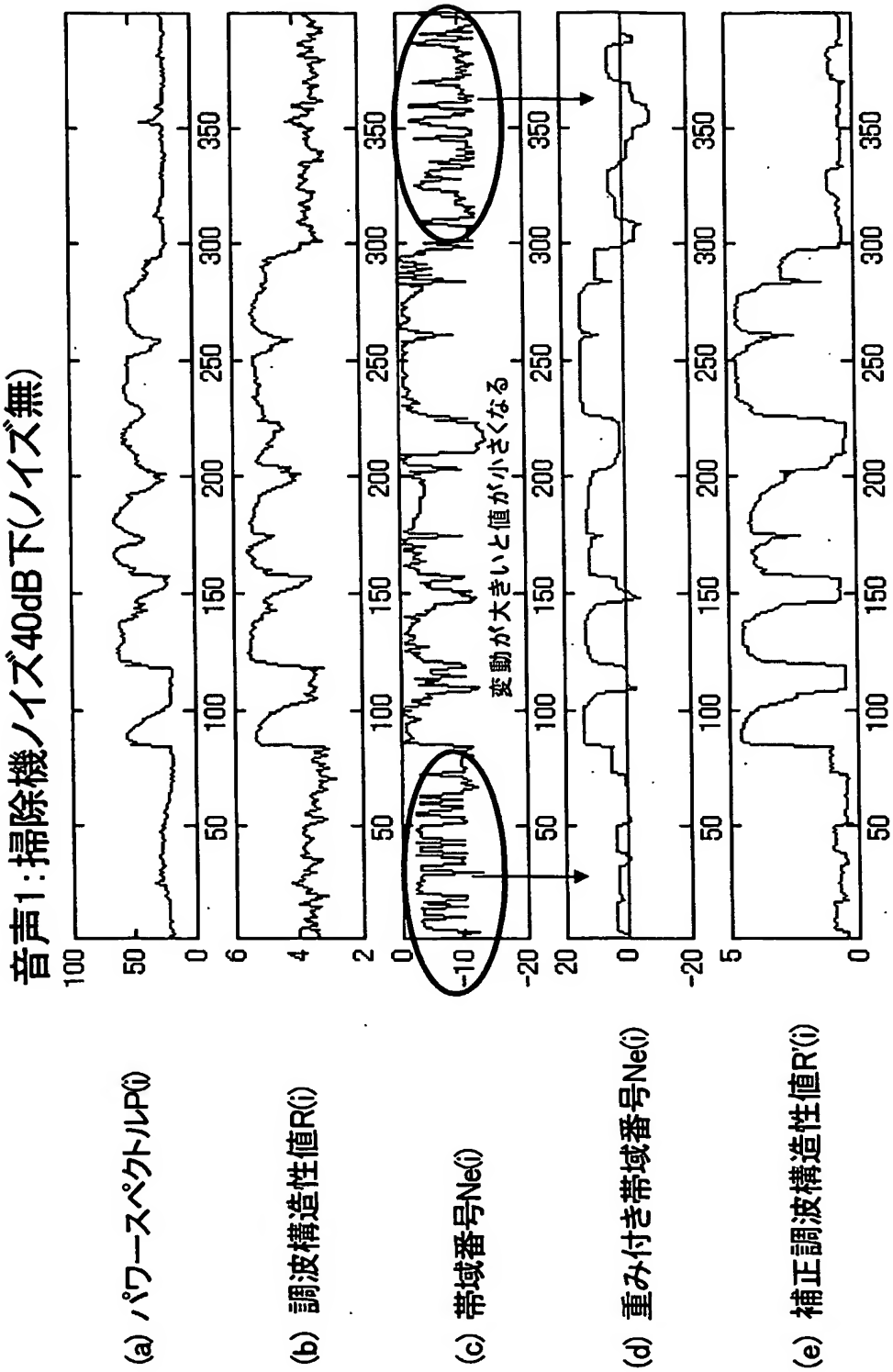

音声1:掃除機ノイズ40dB下(ノイズ無)
(a) パワースペクトルP(i)
(b) 調波構造性値R(i)
(c) 帯域番号Ne(i)
(d) 重み付き帯域番号Ne(i)
(e) 補正調波構造性値R'(i)
変動が大きいと値が小さくなる

図15

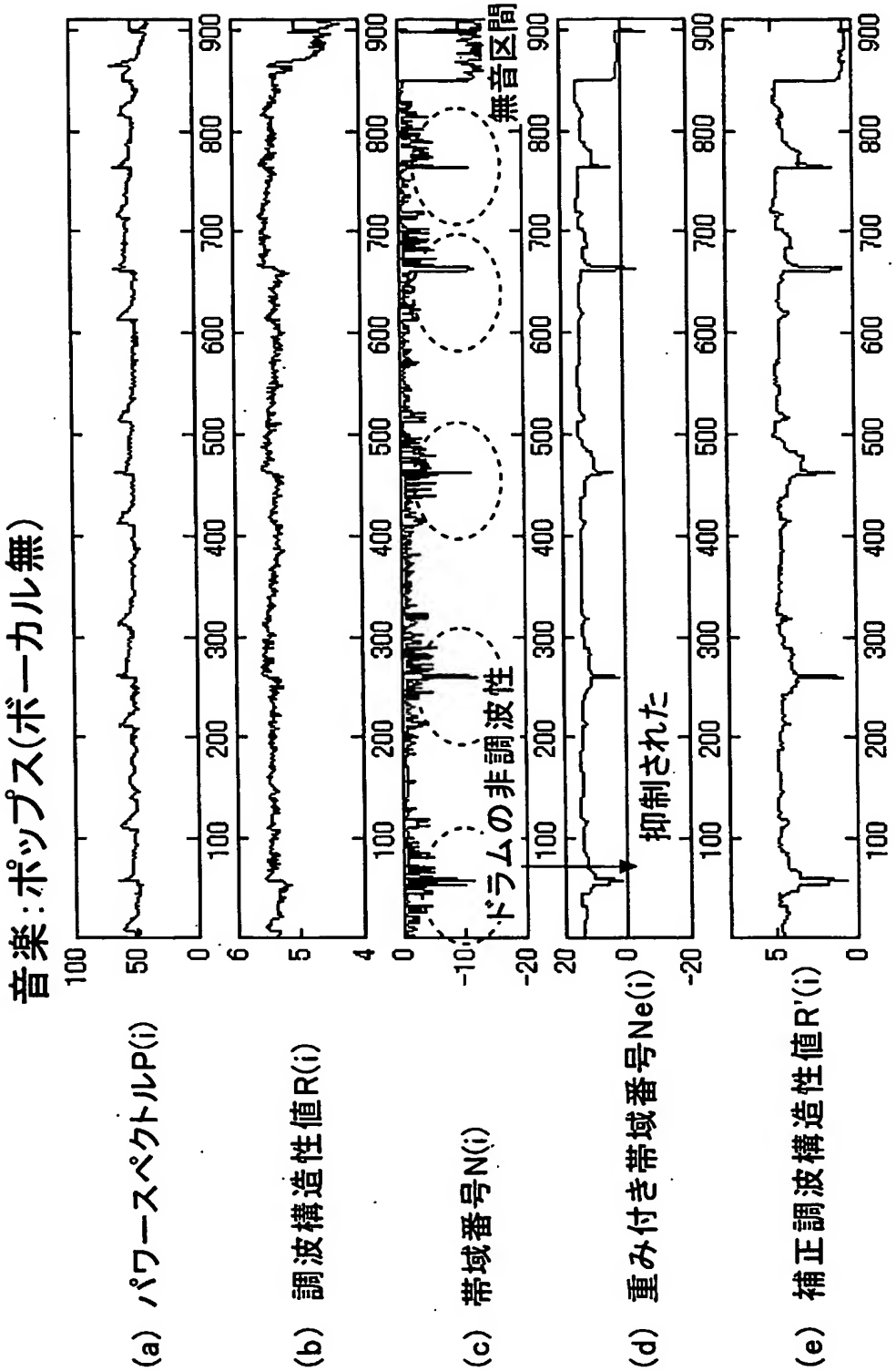

音楽:ポップス(ボーカル無)
(a) パワースペクトルP(i)
(b) 調波構造性値R(i)
(c) 帯域番号N(i)
(d) 重み付き帯域番号Ne(i)
(e) 補正調波構造性値R'(i)
ドラムの非調波性
無音区間
抑制された

図16

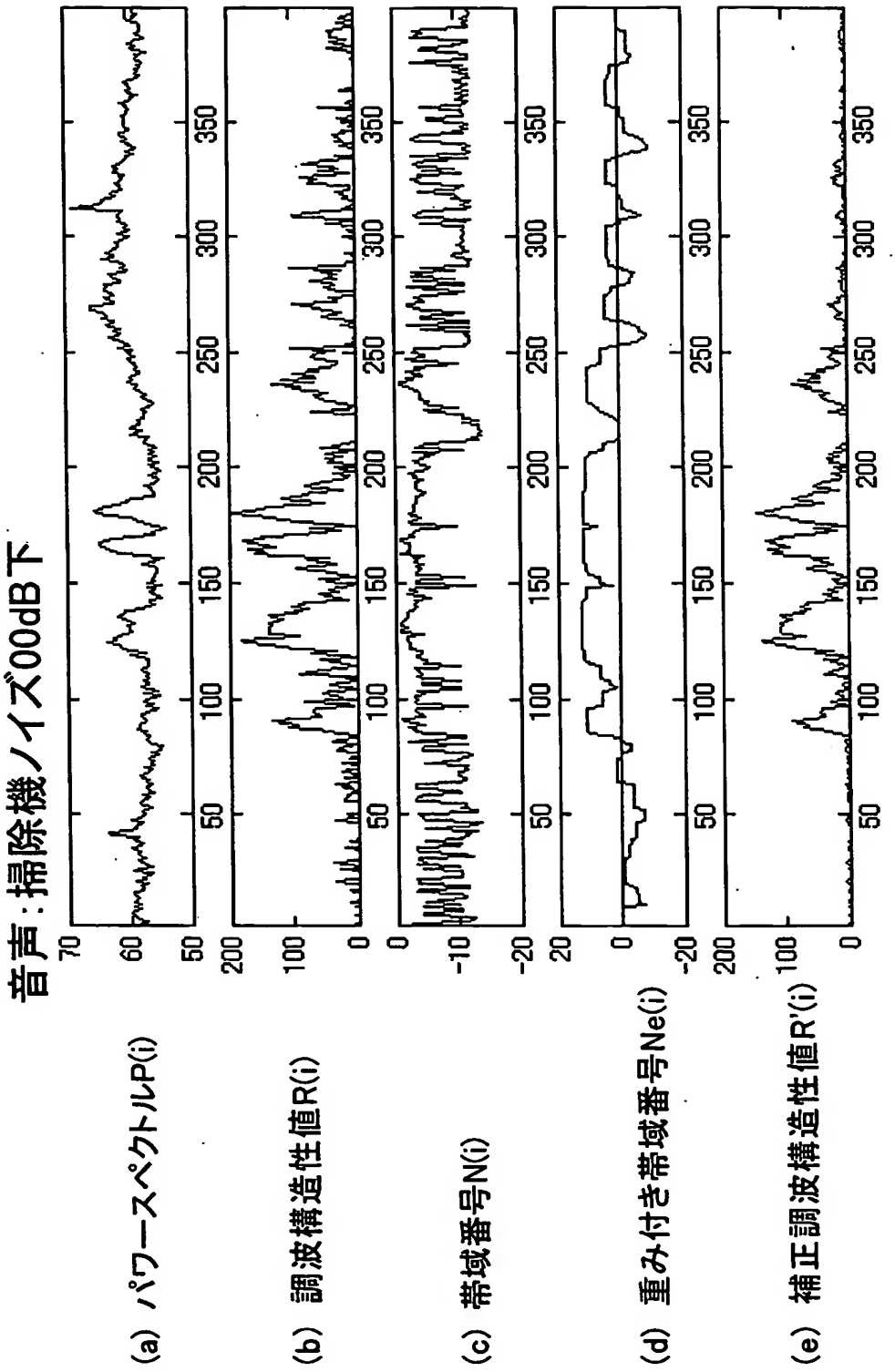

音声:掃除機ノイズ00dB下
(a) パワースペクトルP(i)
(b) 調波構造性値R(i)
(c) 帯域番号N(i)
(d) 重み付き帯域番号Ne(i)
(e) 補正調波構造性値R'(i)

図17

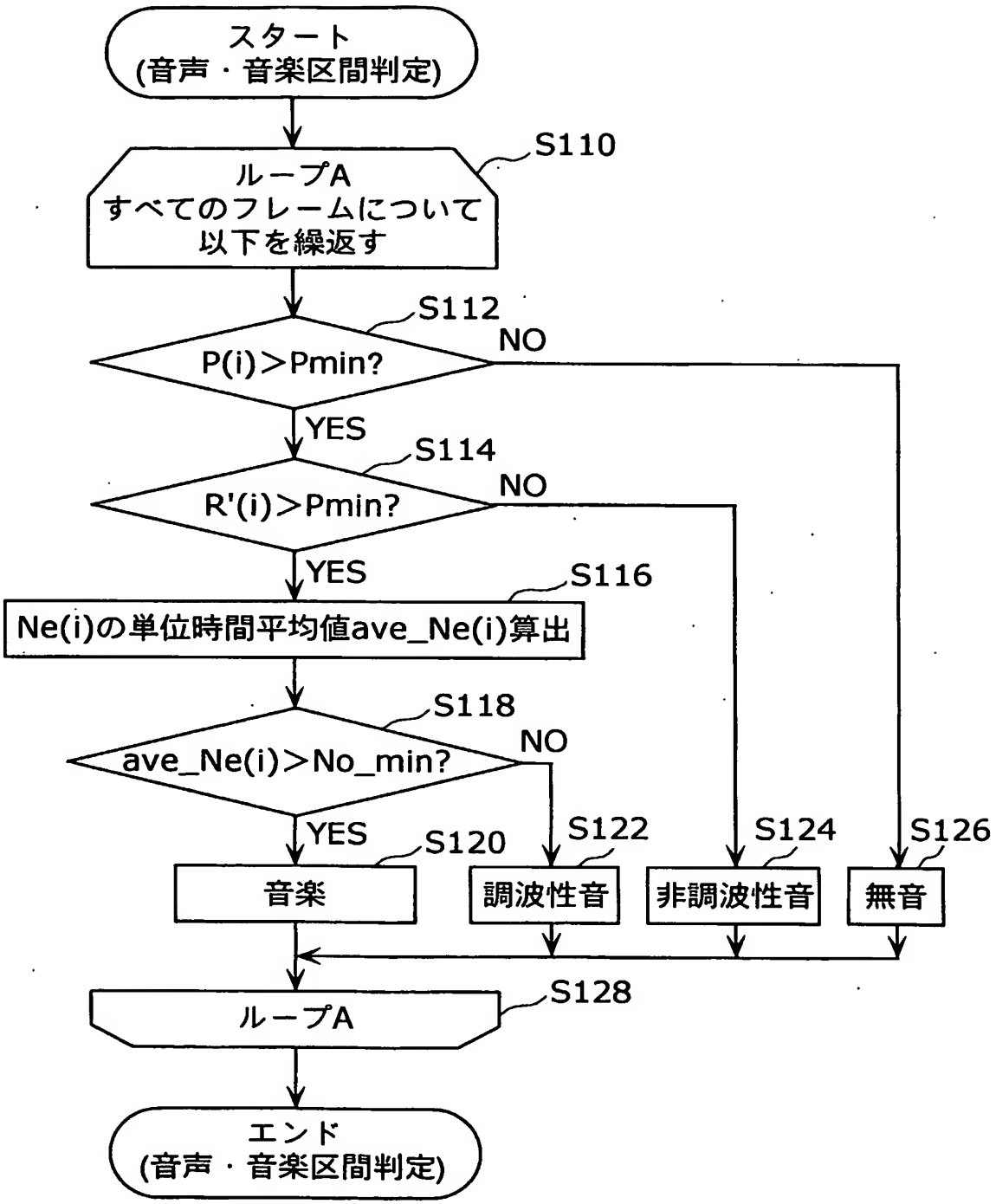

スタート
(音声・音楽区間判定)
S110
ループA
すべてのフレームについて
以下を繰返す
S112
P(i)>Pmin?
NO
YES
S114
R'(i)>Pmin?
NO
YES
S116
Ne(i)の単位時間平均値ave_Ne(i)算出
S118
ave_Ne(i)>No_min?
NO
YES
S120
音楽
S122
調波性音
S124
非調波性音
S126
無音
S128
ループA
エンド
(音声・音楽区間判定)

図18

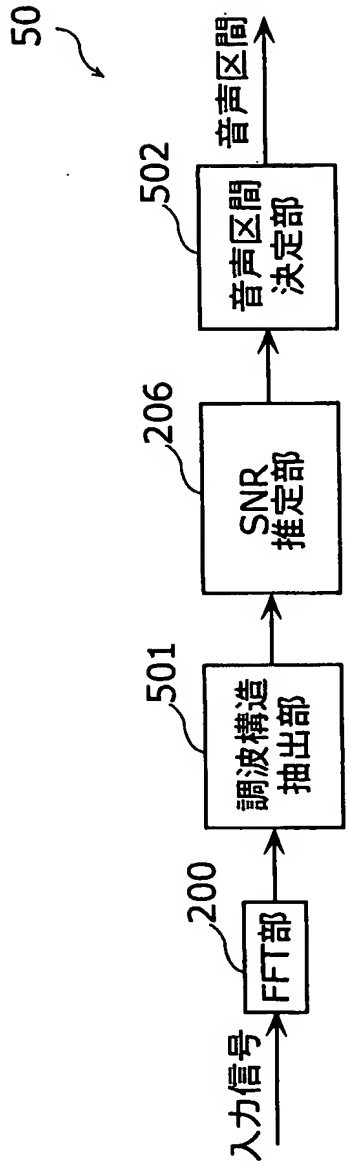


図19

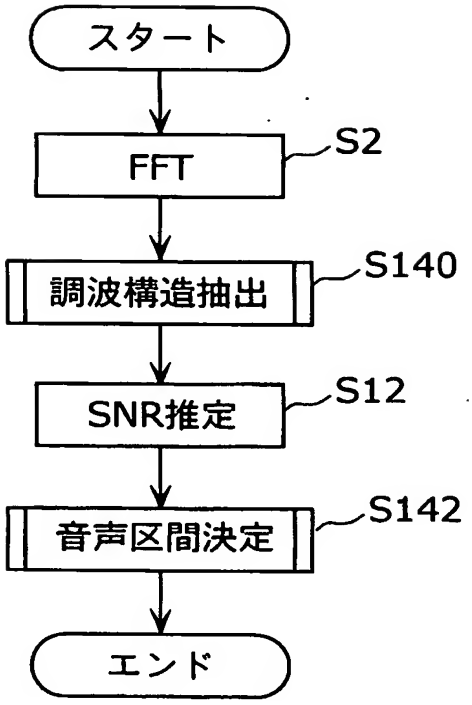


図20

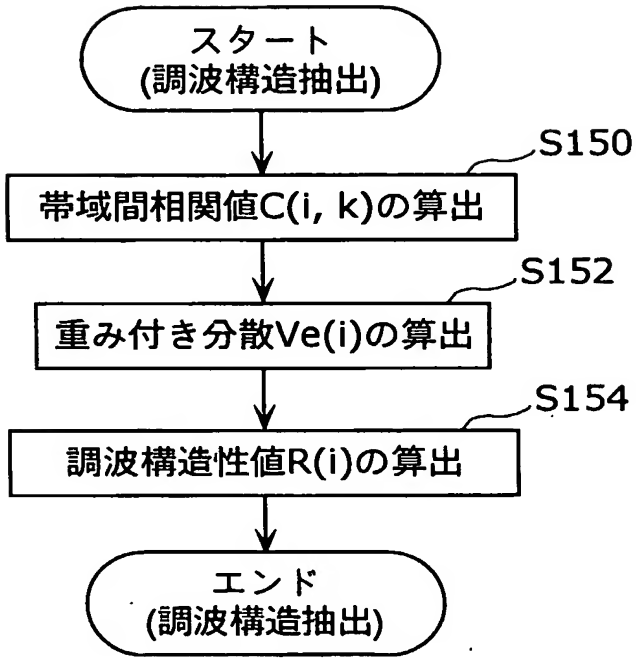

スタート
(調波構造抽出)
S150
帯域間相関値C(i, k)の算出
S152
重み付き分散Ve(i)の算出
S154
調波構造性値R(i)の算出
エンド
(調波構造抽出)

図21

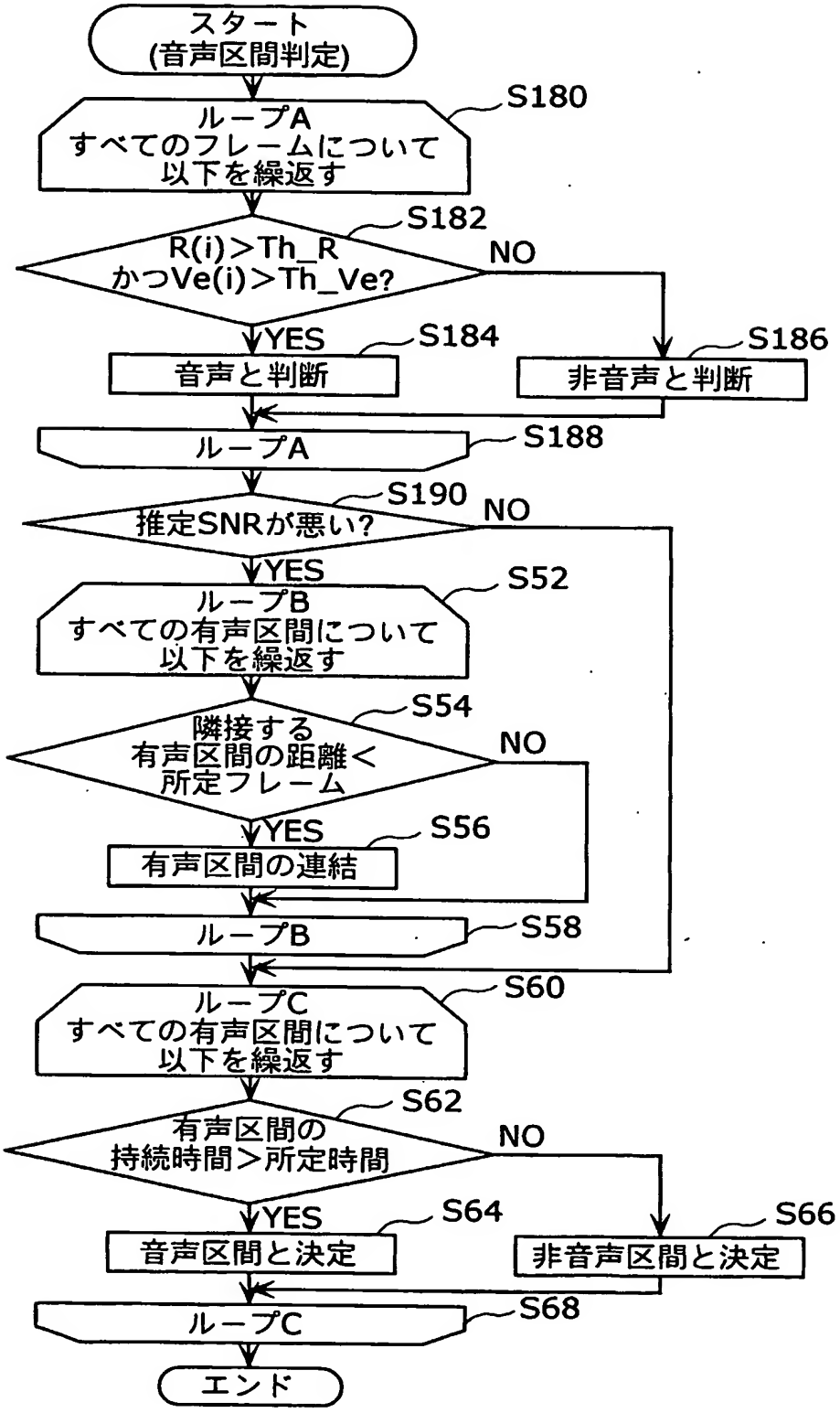

スタート
(音声区間判定)
S180
ループA
すべてのフレームについて
以下を繰返す
S182
R(i)>Th_R
かつVe(i)>Th_Ve?
NO
YES
S184
音声と判断
S186
非音声と判断
ループA
S188
S190
推定SNRが悪い?
NO
YES
S52
ループB
すべての有声区間について
以下を繰返す
S54
隣接する
有声区間の距離<
所定フレーム
NO
YES
S56
有声区間の連結
ループB
S58
S60
ループC
すべての有声区間について
以下を繰返す
S62
有声区間の
持続時間>所定時間
NO
YES
S64
音声区間と決定
S66
非音声区間と決定
ループC
S68
エンド

図22

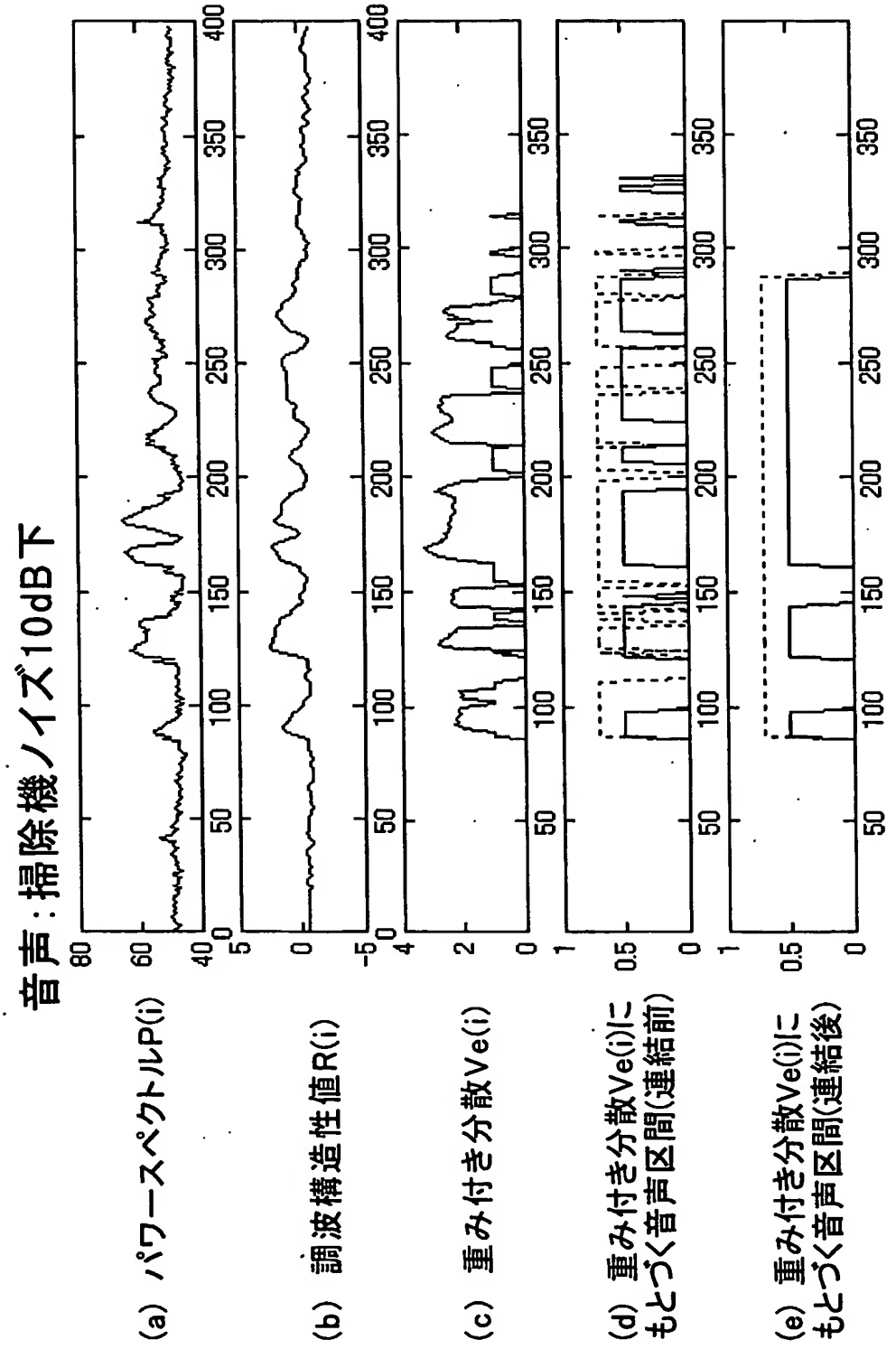

音声:掃除機ノイズ10dB下
(a) パワースペクトルP(i)
(b) 調波構造性値R(i)
(c) 重み付き分散Ve(i)
(d) 重み付き分散Ve(i)にもとづく音声区間(連結前)
(e) 重み付き分散Ve(i)にもとづく音声区間(連結後)

図23

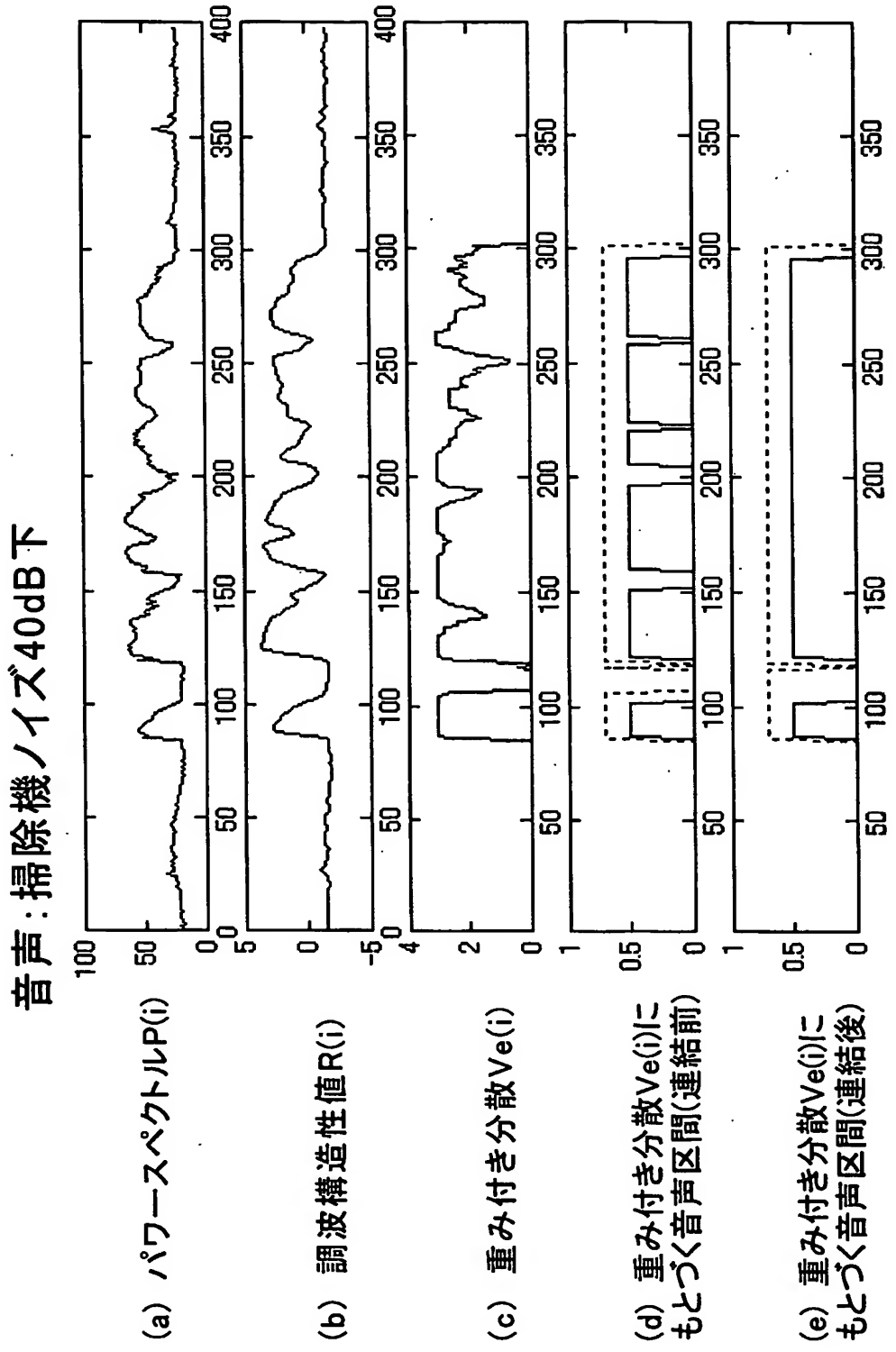

音声：掃除機ノイズ40dB下
(a) パワースペクトルP(i)
(b) 調波構造性値R(i)
(c) 重み付き分散Ve(i)
(d) 重み付き分散Ve(i)に
もとづく音声区間(連結前)
(e) 重み付き分散Ve(i)に
もとづく音声区間(連結後)

図24

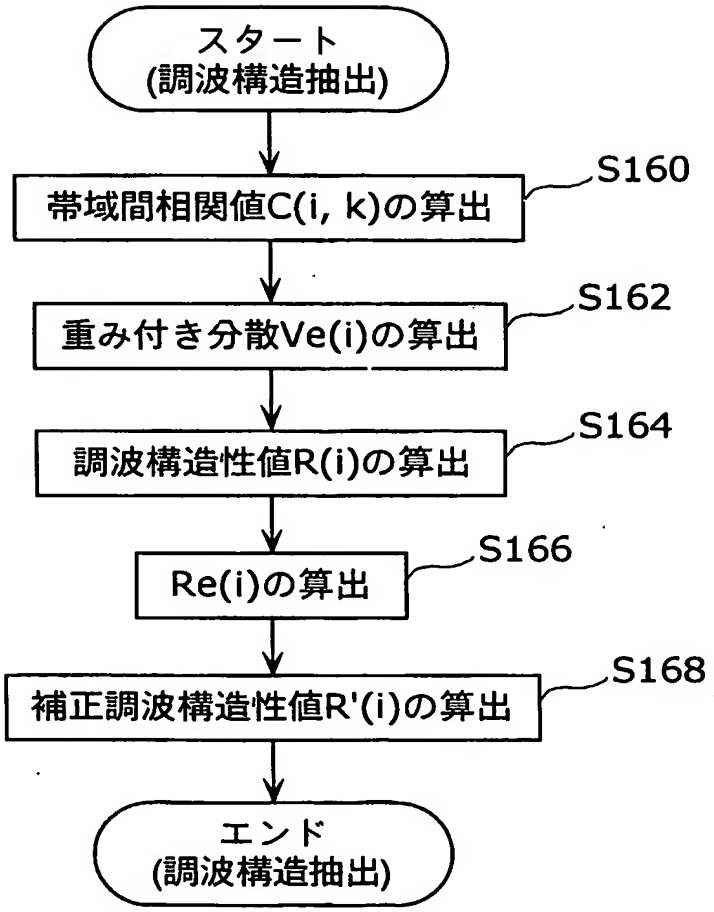

スタート
(調波構造抽出)
帯域間相関値C(i, k)の算出
S160
重み付き分散Ve(i)の算出
S162
調波構造性値R(i)の算出
S164
Re(i)の算出
S166
補正調波構造性値R'(i)の算出
S168
エンド
(調波構造抽出)

図25

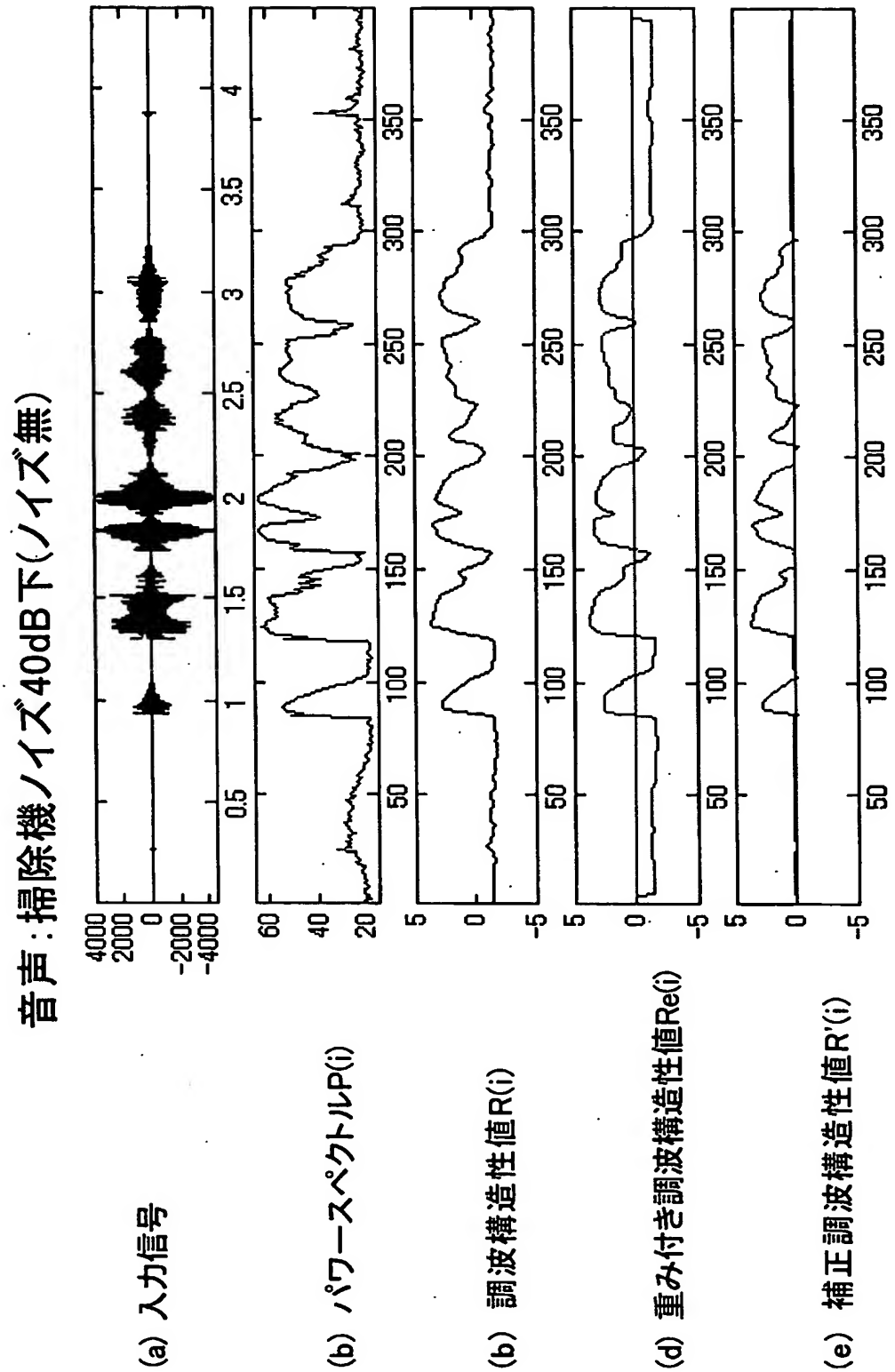

音声：掃除機ノイズ40dB下(ノイズ無)
(a) 入力信号
(b) パワースペクトルP(i)
(b) 調波構造性値R(i)
(d) 重み付き調波構造性値Re(i)
(e) 補正調波構造性値R'(i)

図26

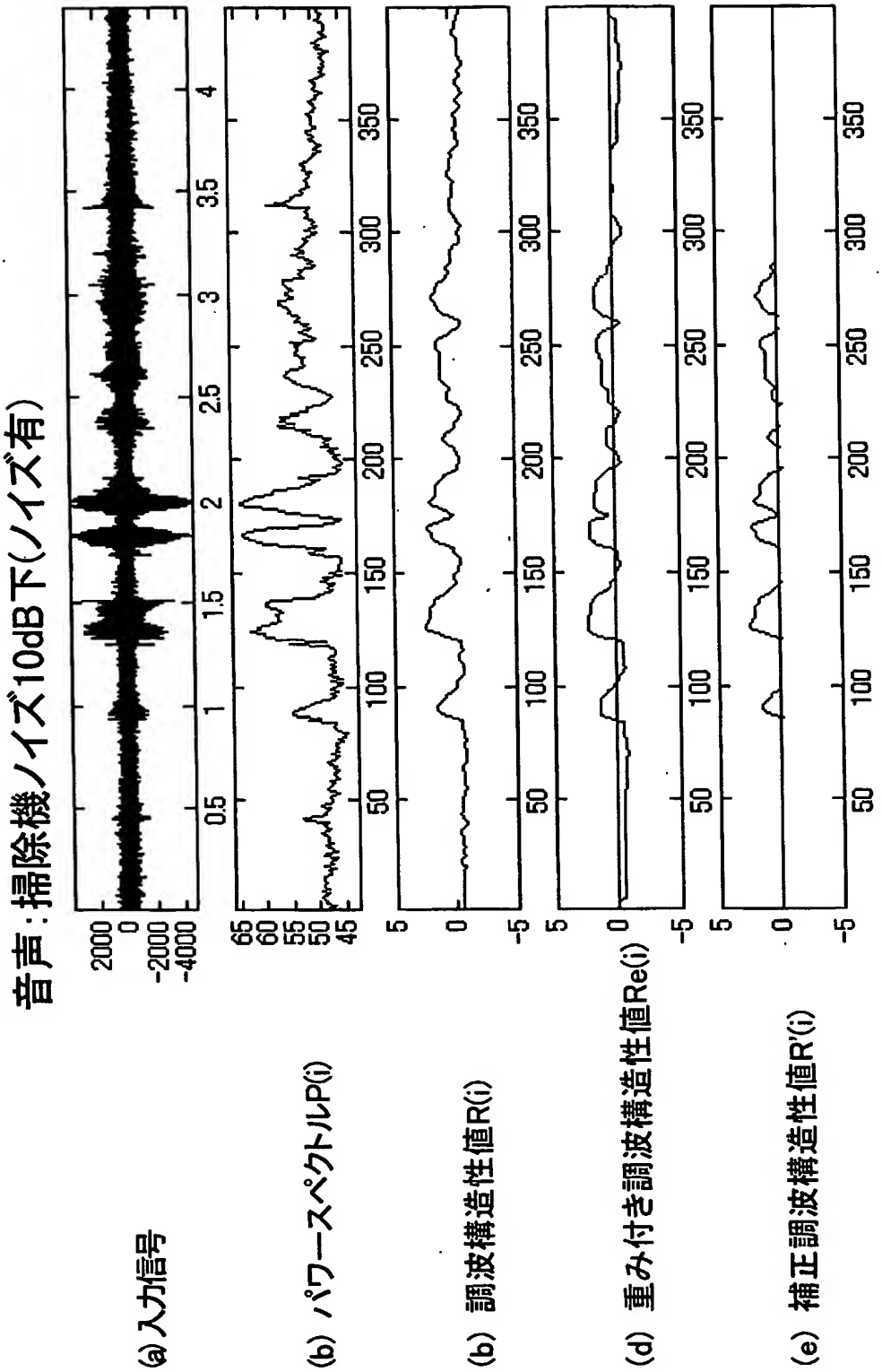

音声:掃除機ノイズ10dB下(ノイズ有)
(a)入力信号
(b) パワースペクトルP(i)
(b) 調波構造性値R(i)
(d) 重み付き調波構造性値Re(i)
(e) 補正調波構造性値R'(i)

図27

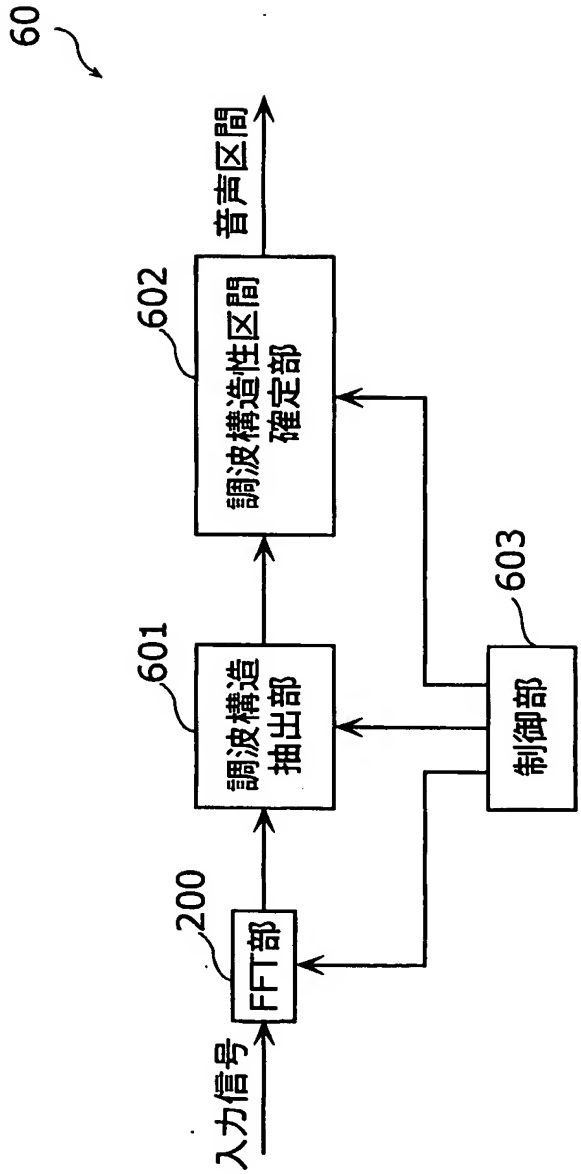


図28

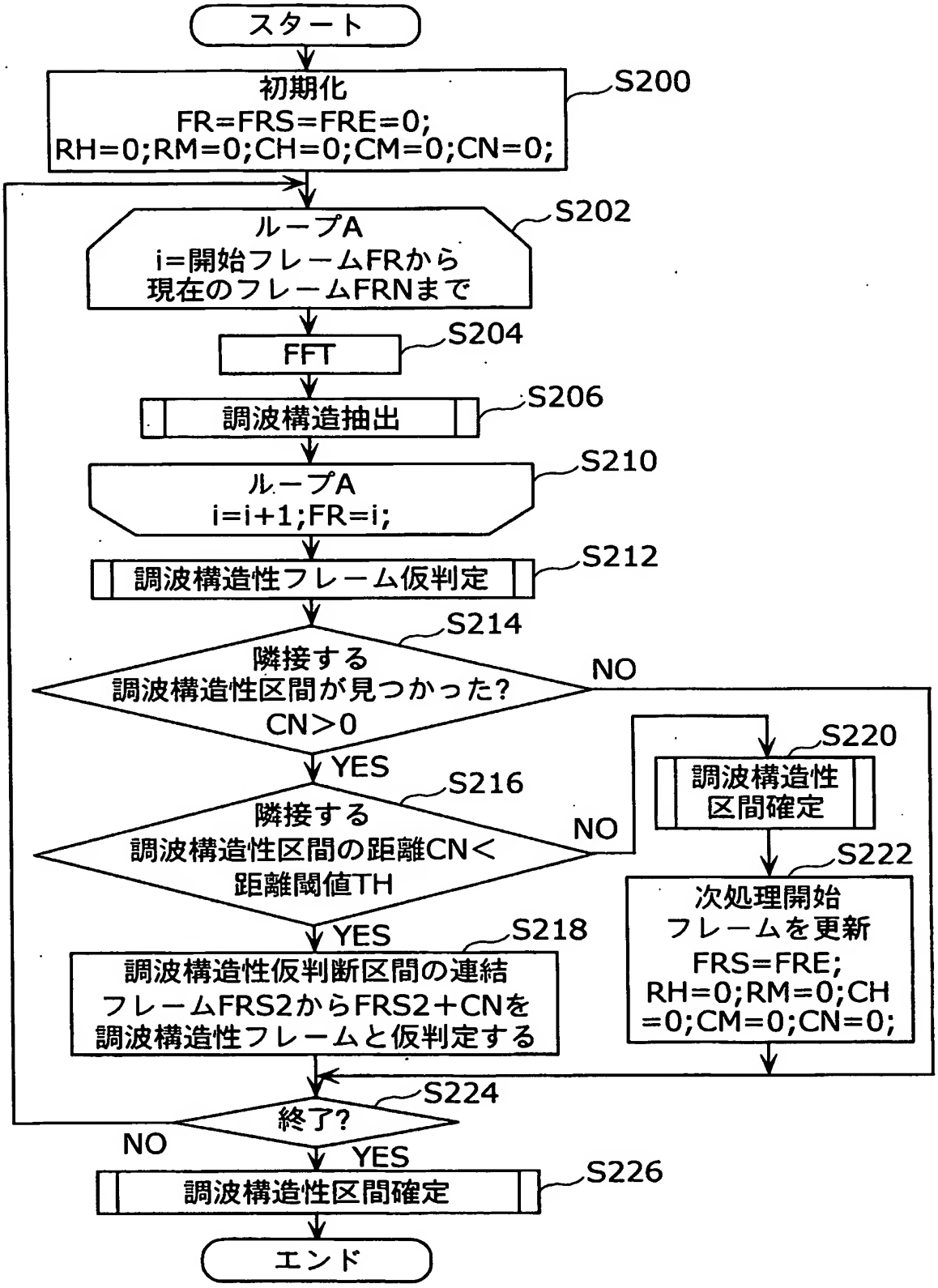

スタート
初期化
FR=FRS=FRE=0;
RH=0;RM=0;CH=0;CM=0;CN=0;
S200
ループA
i=開始フレームFRから
現在のフレームFRNまで
S202
FFT
S204
調波構造抽出
S206
ループA
i=i+1;FR=i;
S210
調波構造性フレーム仮判定
S212
隣接する
調波構造性区間が見つかった?
CN>0
S214
NO
YES
隣接する
調波構造性区間の距離CN<
距離閾値TH
S216
NO
YES
調波構造性
区間確定
S220
次処理開始
フレームを更新
FRS=FRE;
RH=0;RM=0;CH
=0;CM=0;CN=0;
S222
調波構造性仮判断区間の連結
フレームFRS2からFRS2+CNを
調波構造性フレームと仮判定する
S218
終了?
S224
NO
YES
調波構造性区間確定
S226
エンド

図29

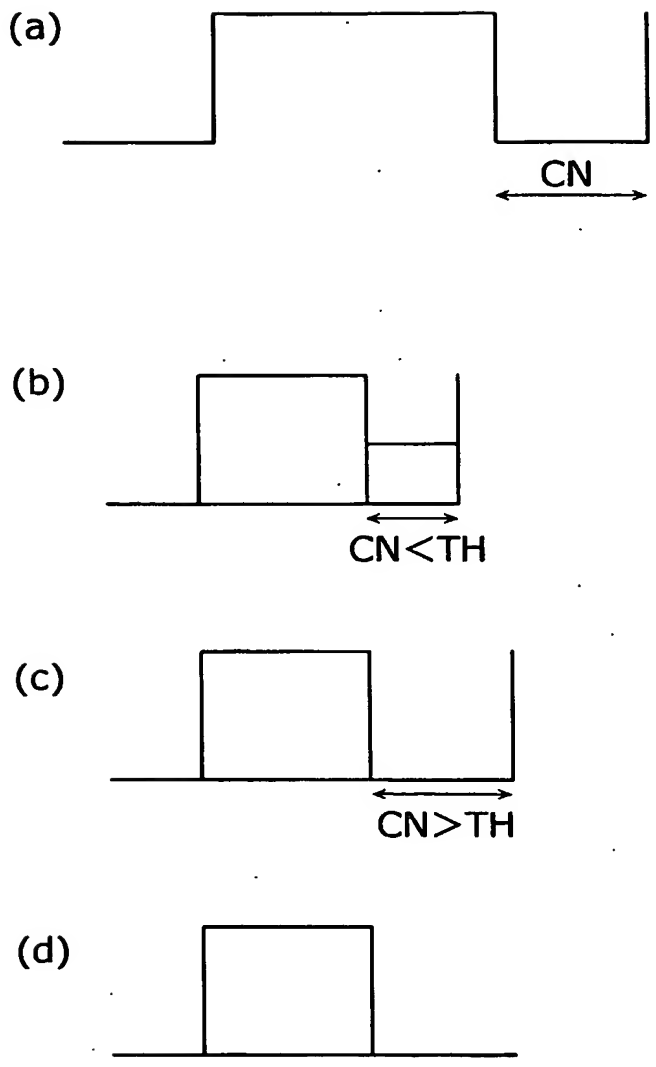


図30

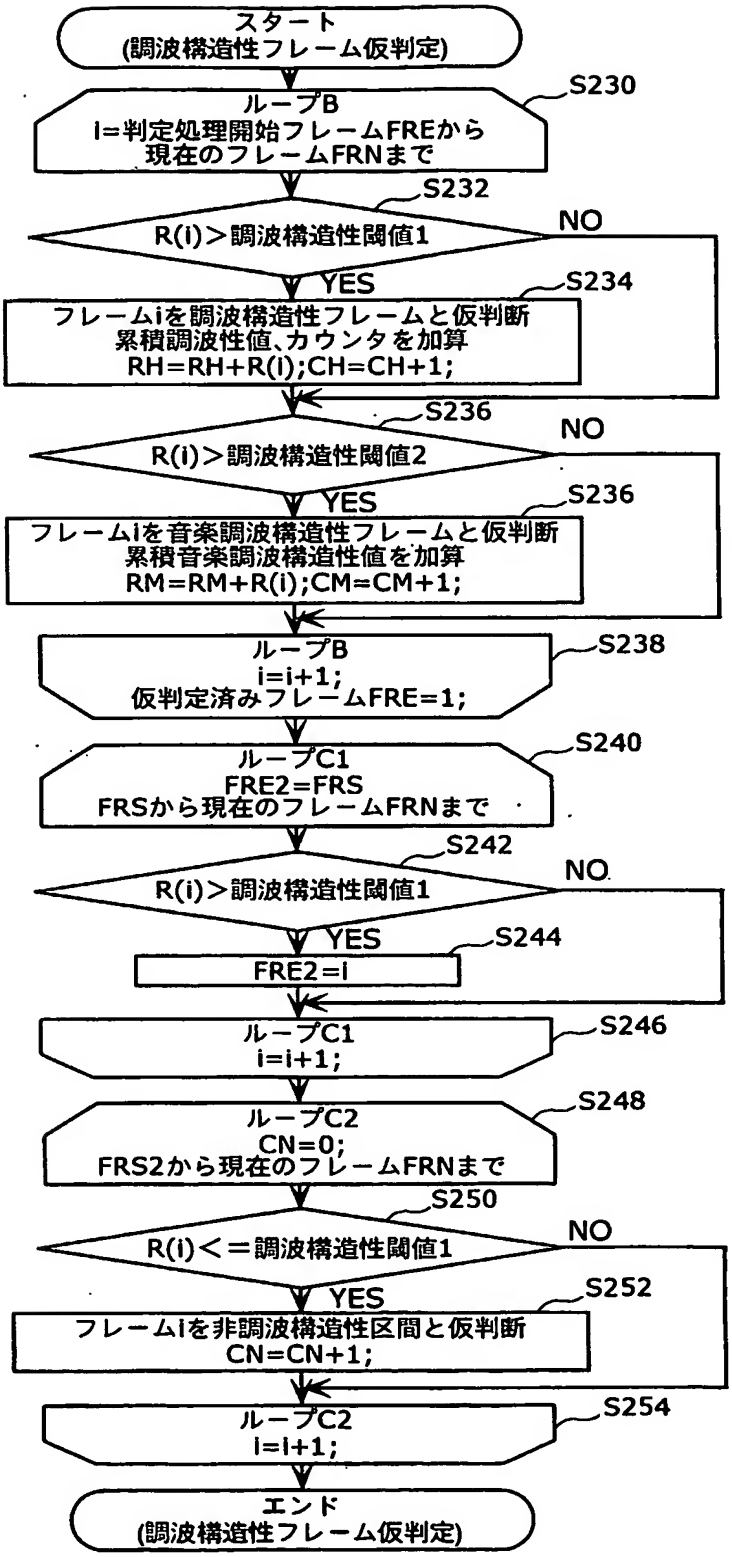

スタート
(調波構造性フレーム仮判定)
S230
ループB
i=判定処理開始フレームFREから
現在のフレームFRNまで
S232
R(i)＞調波構造性閾値1
NO
YES
S234
フレームiを調波構造性フレームと仮判断
累積調波性値、カウンタを加算
RH=RH+R(i);CH=CH+1;
S236
R(i)＞調波構造性閾値2
NO
YES
S236
フレームiを音楽調波構造性フレームと仮判断
累積音楽調波構造性値を加算
RM=RM+R(i);CM=CM+1;
ループB
i=i+1;
仮判定済みフレームFRE=1;
S238
ループC1
FRE2=FRS
FRSから現在のフレームFRNまで
S240
S242
R(i)＞調波構造性閾値1
NO
YES
S244
FRE2=i
ループC1
i=i+1;
S246
ループC2
CN=0;
FRS2から現在のフレームFRNまで
S248
S250
R(i)＜＝調波構造性閾値1
NO
YES
S252
フレームiを非調波構造性区間と仮判断
CN=CN+1;
ループC2
i=i+1;
S254
エンド
(調波構造性フレーム仮判定)

図31

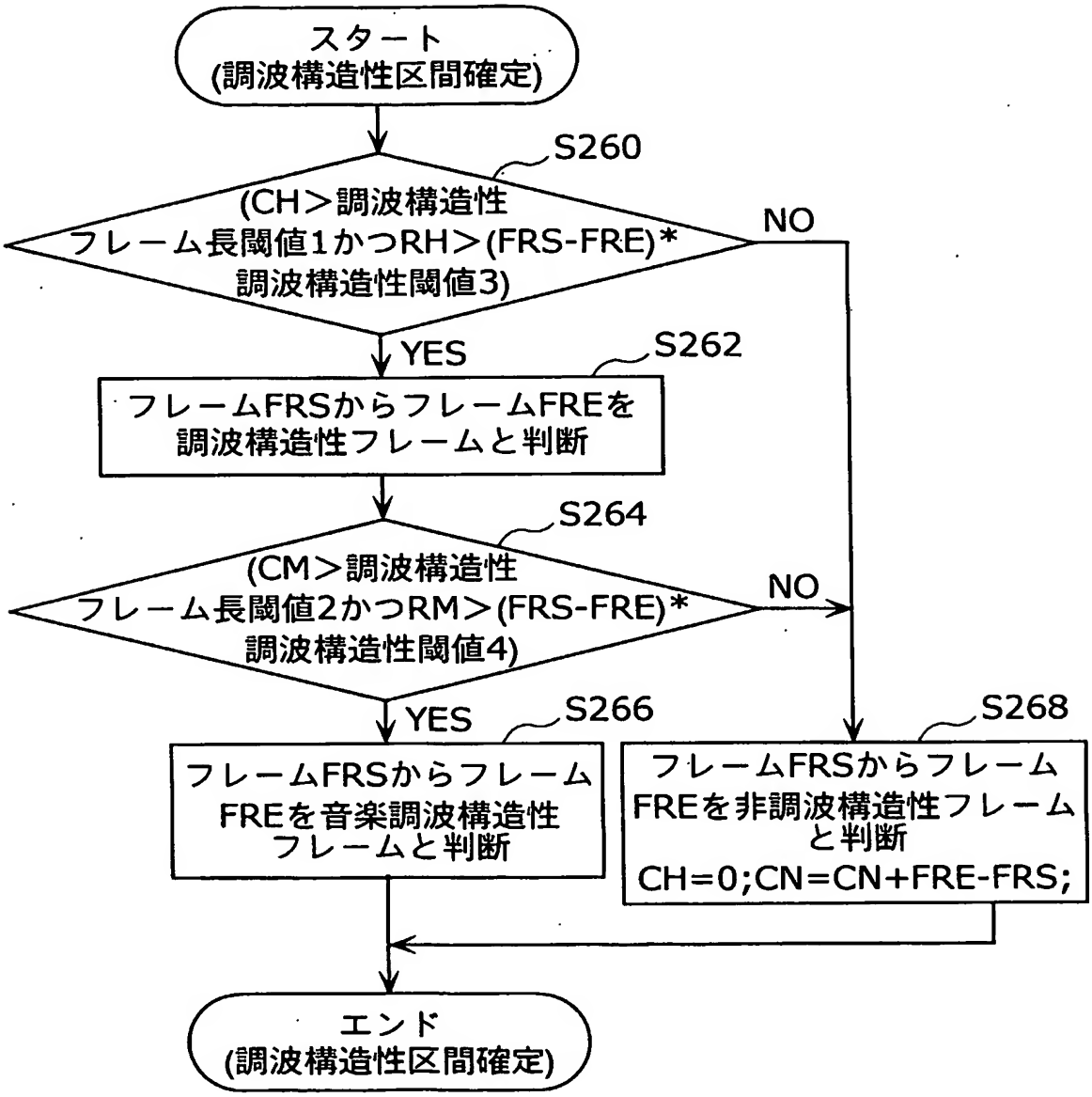
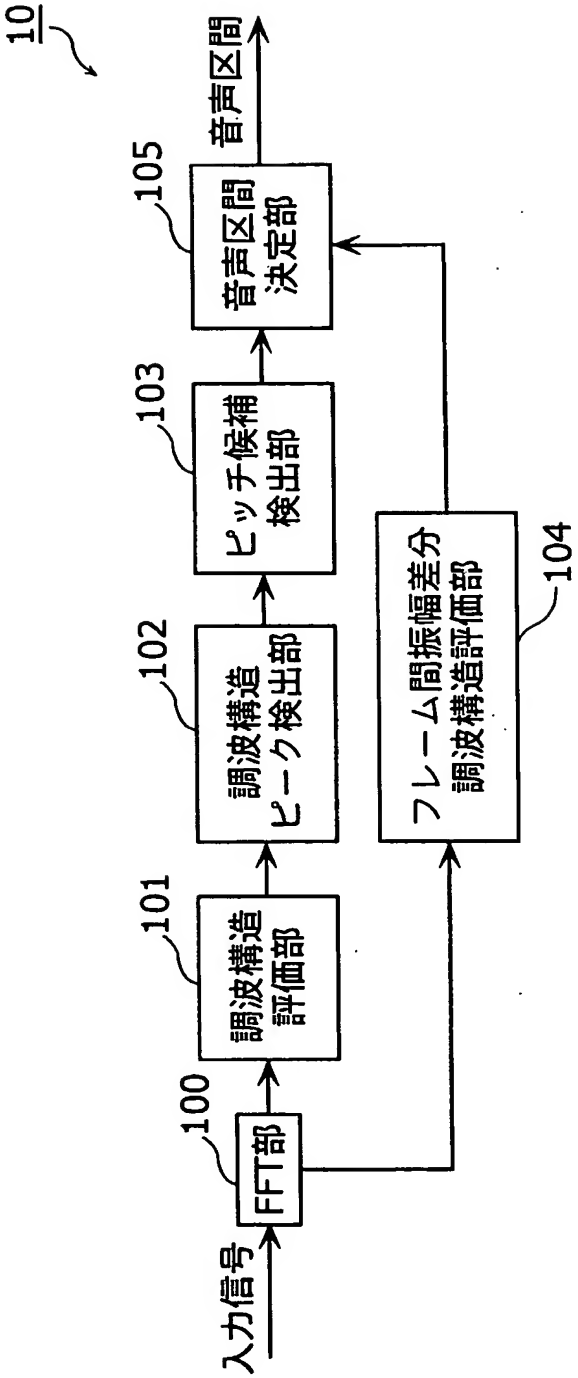

スタート
(調波構造性区間確定)
S260
(CH＞調波構造性
フレーム長閾値1かつRH＞(FRS-FRE)*
調波構造性閾値3)
NO
YES
S262
フレームFRSからフレームFREを
調波構造性フレームと判断
S264
(CM＞調波構造性
フレーム長閾値2かつRM＞(FRS-FRE)*
調波構造性閾値4)
NO
YES
S266
フレームFRSからフレーム
FREを音楽調波構造性
フレームと判断
S268
フレームFRSからフレーム
FREを非調波構造性フレーム
と判断
CH=0;CN=CN+FRE-FRS;
エンド
(調波構造性区間確定)

図32


10
入力信号
FFT部
100
調波構造評価部
101
調波構造ピーク検出部
102
ピッチ候補検出部
103
フレーム間振幅差分調波構造評価部
104
音声区間決定部
105
音声区間

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2004/008051

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl[7] G10L11/02, 06

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
Int.Cl[7] G10L11/02, 06

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Jitsuyo Shinan Koho 1922-1996 Toroku Jitsuyo Shinan Koho 1996-2004
Kokai Jitsuyo Shinan Koho 1971-2004 Jitsuyo Shinan Toroku Koho 1994-2004

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
JSTPlus FILE(JOIS)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2002-162982 A (Matsushita Electric Industrial Co., Ltd.), 07 June, 2002 (07.06.02), Full text; Figs. 1 to 4 (Family: none)	1-17,19-21
A	JP 2001-236085 A (Matsushita Electric Industrial Co., Ltd.), 31 August, 2001 (31.08.01), Full text; Figs. 1 to 15 (Family: none)	1-17,19-21
A	JP 2001-222289 A (Yamaha Corp.), 17 August, 2001 (17.08.01), Full text; Figs. 1 to 7 (Family: none)	1-17,19-21

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 September, 2004 (17.09.04)	05 October, 2004 (05.10.04)
Name and mailing address of the ISA/ Japanese Patent Office	Authorized officer
Facsimile No.	Telephone No.

Form PCT/ISA/210 (second sheet) (January 2004)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2004/008051

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 9-153769 A (Nippon Telegraph And Telephone Corp.), 10 June, 1997 (10.06.97), Full text; Figs. 1 to 6 (Family: none)	1-17,19-21
A	JP 11-143460 A (Nippon Telegraph And Telephone Corp.), 28 May, 1999 (28.05.99), Full text; Figs. 1 to 8 (Family: none)	1-17,19-21

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2004/008051

Box No. II **Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 18
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
Claim 18 describes "a harmonic-structured acoustic signal interval detection device as claimed in claim 28". However, no claim 28 exists in the scope of claims. Accordingly, claim 18 lacks the requirement of clarity within the meaning PCT Article 6.

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III **Observations where unity of invention is lacking (Continuation of item 3 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

国際調査報告

国際出願番号　PCT／JP2004／008051

A.　発明の属する分野の分類（国際特許分類（IPC））

Int. Cl⁷　G10L11／02, 06

B.　調査を行った分野

調査を行った最小限資料（国際特許分類（IPC））

Int. Cl⁷　G10L11／02, 06

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922－1996年
日本国公開実用新案公報	1971－2004年
日本国登録実用新案公報	1996－2004年
日本国実用新案登録公報	1994－2004年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

JSTPlusファイル（JOIS）

C.　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求の範囲の番号
A	JP 2002－162982 A（松下電器産業株式会社）2002.06.07、全文、第1－4図（ファミリーなし）	1-17, 19-21
A	JP 2001－236085 A（松下電器産業株式会社）2001.08.31、全文、第1－15図（ファミリーなし）	1-17, 19-21
A	JP 2001－222289 A（ヤマハ株式会社）2001.08.17、全文、第1－7図（ファミリーなし）	1-17, 19-21

☒ C欄の続きにも文献が列挙されている。　☐ パテントファミリーに関する別紙を参照。

＊　引用文献のカテゴリー
「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日　17.09.2004	国際調査報告の発送日　05.10.2004
国際調査機関の名称及びあて先 日本国特許庁（ISA／JP） 郵便番号100－8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官（権限のある職員）　5C　9381 南　義明 電話番号　03－3581－1101　内線　3541

様式PCT／ISA／210（第2ページ）（2004年1月）

国際調査報告

国際出願番号　PCT／JP2004／008051

C（続き）．　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求の範囲の番号
A	JP 9－153769 A（日本電信電話株式会社） 1997．06．10、全文、第1－6図（ファミリーなし）	1-17, 19-21
A	JP 11－143460 A（日本電信電話株式会社） 1999．05．28、全文、第1－8図（ファミリーなし）	1-17, 19-21

国際調査報告

国際出願番号　ＰＣＴ／ＪＰ２００４／００８０５１

第Ⅱ欄　請求の範囲の一部の調査ができないときの意見（第１ページの２の続き）

法第８条第３項（ＰＣＴ１７条(2)(a)）の規定により、この国際調査報告は次の理由により請求の範囲の一部について作成しなかった。

1．☐ 請求の範囲 _____ は、この国際調査機関が調査をすることを要しない対象に係るものである。
つまり、

2．☒ 請求の範囲 18 は、有意義な国際調査をすることができる程度まで所定の要件を満たしていない国際出願の部分に係るものである。つまり、
「請求の範囲２８項に記載の調波構造性音響信号区間検出装置」と記載する請求項１８は、請求項２８項が請求の範囲に存在しない故、PCT第6条における明確性の要件を欠いている。

3．☐ 請求の範囲 _____ は、従属請求の範囲であってＰＣＴ規則6.4(a)の第２文及び第３文の規定に従って記載されていない。

第Ⅲ欄　発明の単一性が欠如しているときの意見（第１ページの３の続き）

次に述べるようにこの国際出願に二以上の発明があるとこの国際調査機関は認めた。

1．☐ 出願人が必要な追加調査手数料をすべて期間内に納付したので、この国際調査報告は、すべての調査可能な請求の範囲について作成した。

2．☐ 追加調査手数料を要求するまでもなく、すべての調査可能な請求の範囲について調査することができたので、追加調査手数料の納付を求めなかった。

3．☐ 出願人が必要な追加調査手数料を一部のみしか期間内に納付しなかったので、この国際調査報告は、手数料の納付のあった次の請求の範囲のみについて作成した。

4．☐ 出願人が必要な追加調査手数料を期間内に納付しなかったので、この国際調査報告は、請求の範囲の最初に記載されている発明に係る次の請求の範囲について作成した。

追加調査手数料の異議の申立てに関する注意
☐ 追加調査手数料の納付と共に出願人から異議申立てがあった。
☐ 追加調査手数料の納付と共に出願人から異議申立てがなかった。